Computation and representation in emotion and decision-making

Archy Otto de Berker

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Institute of Neurology

UCL

November 2016

I, Archy Otto de Berker, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Archy Otto de Berker, November 2016

Abstract

This thesis deals with three components of an organism's interactions with its environment: learning, decision making, and emotions. In a series of 5 studies, I detail relationships between these processes, and investigate the representation and computations whereby they are achieved. In the first experiment I show how subjective wellbeing is influenced by one's own rewards and expectations, but also those of other people. Furthermore, I find that parameter estimates of empathy predict decision-making in a distinct test of economic generosity. In my second study, I ask how stressful experiences modulate subsequent learning, detailing a specific impairment in action-learning under stress which also manifests itself in altered pupillary responses. In the third, I use a hierarchical model of learning to show that subjective uncertainty in aversive contexts predicts several dimensions of acute stress responses. Furthermore, I find that individuals who show greater uncertainty-tuning in their stress responses are better at predicting the presence of threat. In the final pair of studies I ask how decision variables for value-based choice are represented in the brain. I describe the combination of quality and quantity into value estimates in humans, revealing a central role for the Anterior Cingulate Cortex in value integration using functional magnetic resonance imaging. I next characterize the neural code for value in non-human primate frontal cortex, using single-neuron data from collaborators. These two studies provide convergent evidence that the value code may be more diverse and non-linear than previously reported, potentially conferring the ability to incorporate uncertainty signals directly in the activity of value coding neurons.

Table of contents

Acknowledgements

I have had the pleasure of belonging to two laboratories throughout my time at UCL. Both the Bestmann and Dolan labs have been fantastic places to work, and I'm thankful to all of the people who made each of them so stimulating and convivial. I'd also like to thank the technical staff who made my experiments possible and even pleasurable.

My primary supervisor Sven has been extremely accommodating, encouraging me to pursue my own ideas and supporting me with remarkably consistent good humour. His optimism and courage have provided a steady source of encouragement throughout. Sven taught me to stay close to the data and has emphasized how best to spend my energy and time, neither of which are as infinite as I initially believed. I have felt similarly empowered by Ray, who has advised and supported me both directly and through his unusual ability to assemble an unbelievably talented group of scientists. I feel fortunate to have enjoyed such excellent relationships with not one but two remarkable supervisors.

I have plundered the time of many kind colleagues over the last four years, but none quite so rapaciously as that of Robb Rutledge. With characteristic imperturbability, gentleness, and diligence, he has instructed and inspired at every step of the journey. He has been the first port of call whenever something has gone awry and a source of authority on everything from participant instruction sheets to how to kill a sheep. I am also very grateful indeed for having been involved in his tireless public engagement efforts. Our work at the Roundhouse was one of the highlights of my PhD. Thanks for being such an exemplary mentor and role model.

If Ray and Sven are my scientific fathers and Robb an adored uncle, Zeb Kurth-Nelson has been an older brother. The many hours we have spent together in the lab and outside it have been full of grazed knees, lewd jokes, and finding interesting things under rocks. I have learned a tremendous amount about thinking, science, and thoughtfulness from Zeb, and I look forward to learning more. Any ideas in this thesis that might be considered interesting are probably Zeb's.

I am much indebted to Laurence, Nish, and Steve, who were generous with their data and their time in helping me to understand it. I am in awe of the craft and patience with which they do

science, and the quality of the data that results. It has been a privilege to think about their neurons with them, and to occasionally make cups of coffee for Laurence.

I have made too many friends at UCL to thank them all. However, Federico deserves a special mention as a parmesan importer *par excellence* and a superb housemate. He and Fran have provided the camaraderie and the Michelin-starred lunches necessary to navigate the occasionally rough waters of a doctorate, with Peter battering me on the tennis court whenever I started to get too cocky. When the time came to write the thesis, they were ably succeeded by my parents, who provided the craft beer, homemade bread, and trips to the lido sufficient to get me over the line.

It seems traditional to end by acknowledging the contribution made by one's partner. Steph will take some pride in confessing that she has chiefly strived to thwart my academic ambitions by coaxing me up mountains, into vans, and down ski slopes. I might have got to the end of this degree quicker were it not for Steph; or, more likely, I'd never have got to the end of it at all.

Figures and tables

Chapter 1: General introduction

1.1 Preliminaries

The job of the brain is to find a good mapping between the sensory inputs it receives and the actions it produces [1]. In this chapter we will discuss the reasons why this is a difficult problem, and introduce some of the terminology we need to describe the challenge and the brain's attempts to solve it. We will draw upon efforts in machine learning, where artificial agents tackle analogous problems; from economics, which aims to predict the behaviour of individuals; and from tennis, in which players try to avoid being aced.

The stage in the transformation of perception to action upon which this thesis is focussed is that of value-based decision making. Loosely defined, this is the business of selecting actions that maximise evolutionary fitness, facilitating reproductive success (and its necessary corollary, at least temporary survival). We note that actually estimating the *objective* value in the true sense – the impact upon an organism's rate of gene transmission – is typically impossible. Instead, we borrow from decades of psychology in operationalizing rewards as things that animals will repeat actions to obtain, and punishments as things which animals seek to avoid [2]. By quantifying the relative effort expended in the pursuit and avoidance of different stimuli, we arrive at the notion of *subjective* value, or *utility*, a number we can use to describe how desirable a given outcome is to an organism at a particular moment. There is considerable debate about whether the brain actually represents value [3,4] a question to which we will return. Regardless, value is a very useful concept, as reflected by its prevalence in psychology, economics, and computer science [5,6].

This thesis is concerned with three interlinked processes by which organisms interact with valuable stimuli: learning, decision-making, and emotion (Figure 1.1). We start with a brief justification of this characterisation, before touching upon the algorithms and neurobiology thought to underpin each of these processes, highlighting the difference between approaches derived from reinforcement learning and from a Bayesian perspective and discussing how each has contributed to our understanding of the brain.

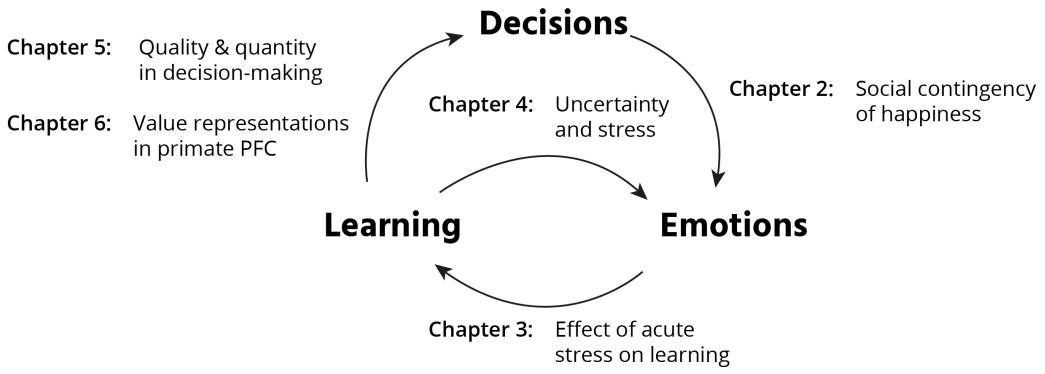


Figure 1.1 | Framework for this thesis An organism seeking to maximise value takes decisions, learns from the results, and experiences emotion along the way. The experiments presented here pertain to the relationships between these processes.

1.1.1 Deciding, learning, feeling

Figure 1.1 shows how the experiments in this thesis concern the interplay of learning, decisions, and emotional processes. It contains several simplifications. The first is the neat cyclical nature, which implies a repetitive serial process. In reality, the relationships between these three elements are probably bidirectional and dynamic. The arrows I have chosen reflect the topics I will deal with here, and omit, for instance, the influence of current emotional state upon decision-making, which is well-documented [7]. Secondly, the use of single arrows understates the degree to which multiple learning and decisional systems act in concert. As we shall see, there is abundant evidence for multiple learning mechanisms in the brain, which provide inputs to decisional-processes which evolve in parallel in numerous frames of reference [8]. Whether there are similarly numerous links between decisional and emotional systems remains to be seen.

To what extent is each of these components necessary? Learning is clearly essential if an organism is to interact successfully with the world; even plants are endowed with some ability to modulate their behaviour on the basis of experience [9]. But why do we need a separate system for making decisions? Impressively capable neural networks do not display a clear segregation of learning and decisional systems, instead learning a direct mapping from inputs to actions [10,11]. One conceptual reason why it might be useful to think about decisions as being somewhat

separate from learning is alluded to above; we have multiple learning systems, which advocate distinct courses of actions when confronted with a given decision. A distinct system for making decisions is able to draw upon numerous sources of evidence, knitting together the outputs of different learning systems to produce a decision [12]. Furthermore, such a system can be used to flexibly link up inputs and outputs, so that knowledge about the climate of different Latin American countries can be combined with experience of different confectionary brands to allow us to choose a particular bar of chocolate, or integrated with airline prices to decide where to go on holiday.

What is less clear is why learning and decision-making need be associated with fluctuations in emotional state, a problem intimately related to that of consciousness itself. Such philosophy is beyond the scope of this thesis, but we hope that by delineating the properties and features of the relationship between emotions and the environment we might glean something of their function [13]. Towards the end of this chapter, we highlight recent attempts to understand the function of emotion through a characterization of its causes and consequences.

1.2 Value and decisions

In this section we will briefly expand upon the notions of value and utility, and the way that humans combine information to make choices. In doing so, we will define the target for learning – the information we think that organisms should acquire in order to make good decisions. Much of the language of this section will come from economics, which seeks to predict peoples' choices by estimating the subjective value associated with different options; in the following sections we shall turn to the lexis of machine and statistical learning, which describe how such values might be assigned.

1.2.1 Economics and decision-making

Given the central role of money in society, it is unsurprising that the first concerted efforts to understand valuation focused upon financial decisions. The starting point was on a distinctly un-biological question: given an offer of a gamble of specified magnitude and probability, how did people behave? Despite the seemingly abstruse relationship to the problems that the brain evolved to solve, answering this question has proved surprisingly occupying in both economic

and biological sciences. Blaise Pascal in the 17th century defined an Expected Value (EV) term which described a strictly linear relationship between objective and subjective value:

$$EV = Probability * Magnitude$$

Equation 1.1

However, this linear description was declared insufficient by Daniel Bernoulli in 1738, who noted that 'a ducat is worth more to a pauper than a Prince'. Bernoulli borrowed a problem from his brother, Nicolas, to illustrate this point. The 'St Petersburg' paradox describes a hypothetical gamble in which a coin is repeatedly flipped. The pot starts at two ducats. Each time the coin comes up heads, we double the amount in the pot. When the coin eventually comes up tails, you receive the quantity in the pot. How many ducats would you pay to play this game?

According a Pascallian notion of value, we should be happy to pay *any* amount. This is because the EV of the series is the weighted sum of the probabilities of each outcome. We therefore sum to the point of a vanishingly small chance of obtaining an astronomical large amount of money:

$$EV = \frac{1}{2} * 2 + \frac{1}{4} * 4 + \frac{1}{8} * 8 \ldots \sim 0 * \sim \infty$$

Equation 1.2

Resulting in an infinite sum. Although laboratory experiments on behaviour in infinite St Petersburg games is sparse (possibly because the prospect of having to pay participants an infinite sum does not wash with investigators clutching hard-won grants) finite-sum versions of the game illustrate that people are risk-averse, paying far less than they 'ought' to part with money in exchange for small chances of great sums of money [14]. This is consistent with Bernoulli's own conclusion, that paupers value ducats more than princes, suggesting that the relationship between subjective utility and value is concave; increasing gains in wealth return decreasing increases in utility. This idea was developed further by Kahneman and Tversky, who described a series of deviations from utility theory which could be accounted for by modifying the utility function such that it was concave for gains but convex for losses, explaining the observation that people appear risk-averse for gains but risk-seeking for losses [5], whilst acknowledging non-linear assessments of probability and the power of contextual effects upon choice [15].

Adding to the nuanced account of choice provided by Kahneman, Tversky and others [16] was the observation that people frequently made choices along axes unpredicted by a purely economic model. People possess a variety of well-characterized social preferences which violate economic dicta, such as for fairness and equality [17,18] [19-21]. Monkeys display similar preferences [22]. Indeed, some have proposed that such preferences are essential for the formation of large societies [23]. Recent work has identified the neural basis of socially-derived value (see below and [24]) and, as we report in **Chapter 2**, preferences for equality are also expressed in fluctuations in participants' emotional state.

By the end of 20th century, therefore, economics and psychology had established that people possessed internal value functions that were consistently different from those prescribed by probability alone; that attitudes towards gains were not symmetric with those towards losses; and that, in a variety of situations, peoples' valuations were sensitive to contextual and framing effects.

1.2.2 Neural representations of value

The economic dissection of value presented above poses an obvious question: is this variable, subjective utility, represented in the brain? Somewhat unsurprisingly, the answer is yes. Setting aside for a second the thorny issue of discriminating between representations of value and its correlates [25], it is clear that when presented with different options, neural responses in a variety of brain regions [26] tend to reflect subjective utility rather than objective Pascallian value [27-29].

Pausing at this point, we might wish to bring the discussion back onto more ethological ground. Is the utility so painstakingly defined by economists the same metric we use to make decisions about biologically relevant variables such as food, water, and sex? The answer is probably yes, with caveats. Several authors have observed that to make decisions between *incommensurable objects*, we need a common scale along which different kinds of reward can be compared [30] [4,31]. Early work identified the ventromedial prefrontal cortex (vmPFC), ventral striatum, and orbitofrontal cortex (OFC) as the most promising locus of an abstract value signal [30] [26,31], although more detailed attempts to parse the stimulus-specificity of signals in each region suggest that the OFC value signal may be stimulus-dependent [32]. More abstract stimuli such as the value of a humorous joke [33] or a beautiful face [34] also appear to activate this network of

brain regions, as do modulations of value by social context [35]. Parallel work in non-human primates has described single cells in the PFC and elsewhere that integrate information about probability, magnitude, effort, and delay in a manner suggestive of a subjective utility signal [36-39]. The prevailing view over the last decade, as captured by the term 'neuroeconomics', is thus that the brain does indeed calculate values for different offers in a common metric [40,41]. As we shall see, this aligns very well with accounts of reinforcement learning, which suggest that agents should store value estimates associated with certain states, and use these value estimates to decide what to do.

1.2.3 Mechanisms for making value-based decisions

We have only touched upon the mechanisms whereby animals actually *make* decisions. It may seem natural from the account above to suppose that the brain calculates the value of different options, and then compares them. This has been the dominant viewpoint in economics, and by extension neuroeconomics, for some time [40]. However, recent work suggests that this view is too simplistic, at least when describing choices involving the integration of multiple pieces of information. When comparing options that differ in both probability and magnitude, behavioural and neural evidence suggest that people compare options on a by-attribute basis, before making a choice that is biased towards the attribute that is the most discriminative {hunt:2014gq}. For example, house-buying choices proceed as a comparison of prices, size, neighbourhood crime statistics etc., rather than by a summation over all of these terms followed by a comparison step [42]. This finding, underpinned by a neural network model of competition through mutual inhibition [43,44], suggests that the ingredients of a choice are more important than previously acknowledged, at least in neuroscience [3]. A complementary ethological perspective emphasises the continuous nature of action-selection, suggesting that serial processing models invoking valuation stages followed by choice and action provide an impoverished perspective on decision-making [8].

This also poses a question to which we will return in due course: whilst economics posits that a single representation of value is sufficient to underpin all value-guided choice, neural perspectives emphasise that representations should be subservient to computations [45,46]. To solve a task where you're estimating how much you'd be happy to pay for a food item, using an integrated value representation might be effective [30], but choices between complicated offers

differing along multiple dimensions might be more effectively represented using an attribute-based code[44]. Finally, and as we shall see, the value estimates used to compare options might be simultaneously sampled from multiple learning systems. Whilst economics places great emphasis upon consistency, the brain suffers from no such constraints, marching instead to the drum of evolution. From this perspective, we might expect regions of the brain performing different value-based *computations* to harbour distinct *representations* of value. Having planted the seeds of this idea, we will leave them until **Chapters 5** and **6** to germinate.

1.2.4 Summary and relevance

Economic approaches to understanding choice suggest the existence of a value function, along which options are compared. The decisions made by humans appear to obey some internal estimate of value, rather than that prescribed by probability theory. Neural correlates of this internal estimate are commonly identified in the brain, particularly in the prefrontal cortices and basal ganglia. However, we emphasise that not all value-based decisions are made on the basis of an integrated value signal, a subject to which we will return in later sections. We now turn to the question of how organisms come to associate value with the objects and actions between which they make decisions; how an individual comes to prefer an apple to an orange, or a top-spin backhand to a slice.

Several experiments in this thesis rely upon the notion of expected value and utility. In **Chapter 2** I present an experiment in which subjects make choices between certain outcomes and gambles, which we document in terms of their expected value. In **Chapter 5** I present a brain imaging study examining the integration of quality and quantity in the formation of value signals, using economic tools to estimate the quality of different stimuli, and in **Chapter 6** I analyse single-neuron responses to cues of different probability and magnitude in an attempt to characterise the representation of value in a variety of prefrontal brain regions.

1.3 Learning from experience

Bernoulli's prince and pauper both assign value to ducats, and the common market determines the value of everything else in terms of ducats. For most of history, assigning value to objects has not been so straightforward. A monkey coming to a new tree festooned with an unfamiliar fruit must *learn* the value of this new fruit, to be able to compare it to her old favourites. As we

shall see, monkeys and other animals are thought to rely upon several different algorithms to achieve this.

1.3.1 Predicting rewards

We spend a great deal of time pursuing things that are not intrinsically (biologically) rewarding. Stimuli that act as surrogate rewards are known as *secondary rewards*, to distinguish them from the primary rewards that have direct biological relevance. Money, bicycles, and books all come into this category. Learning to value things that are *predictive* of reward appears to be a powerful mechanism for eventually attaining it.

Pavlov's classic experiments demonstrated that animals are highly sensitive to the contingency between neutral and valuable stimuli. Pavlovian, or *classical,* conditioning describes the association between a conditioned stimulus (CS), and an unconditioned stimulus (US). Over time, the physiological responses associated with the latter come to be elicited by the former. These can include paying more attention to the item in question, approaching it, or even attempting to eat it. This process inspired the hugely influential Rescorla-Wagner model, which captures several fundamental aspects of learning in humans and other animals.

1.3.1.1 The Rescorla-Wagner model of conditioning

The Rescorla-Wagner (RW) model describes how classical conditioning occurs. It rests upon the idea that animals learn when an experience deviates from their expectations; that is, it is *surprising*. This idea has had a profound influence upon our understanding of learning, with a legacy that reverberates from electrophysiology to artificial intelligence. Formally, the RW model describes the incremental increase in strength of an association through *prediction errors*:

$$\Delta V_A = \alpha\beta(\lambda - \sum V_{AX})$$

$$V_{A,t+1} = V_{A,t} + \Delta V_A$$

Equation 1.3

Where ΔV_A describes the change in associative strength for stimulus A, λ is the maximum associative strength, and $\sum V_{AX}$ is the current associative strength of A and all the other stimuli in the environment (lumped together as X). In words, the Rescorla-Wagner rule tells us that associations are driven by the difference between the true association between CS and US and

the animal's current expectation of that association. This produces a learning curve which decreases in slope as learning proceeds (Figure 1.2). The terms α and β control the rate of acquisition, and are usually interpreted to be properties of the stimulus, although they can also reflect knowledge of the environment's volatility [47].

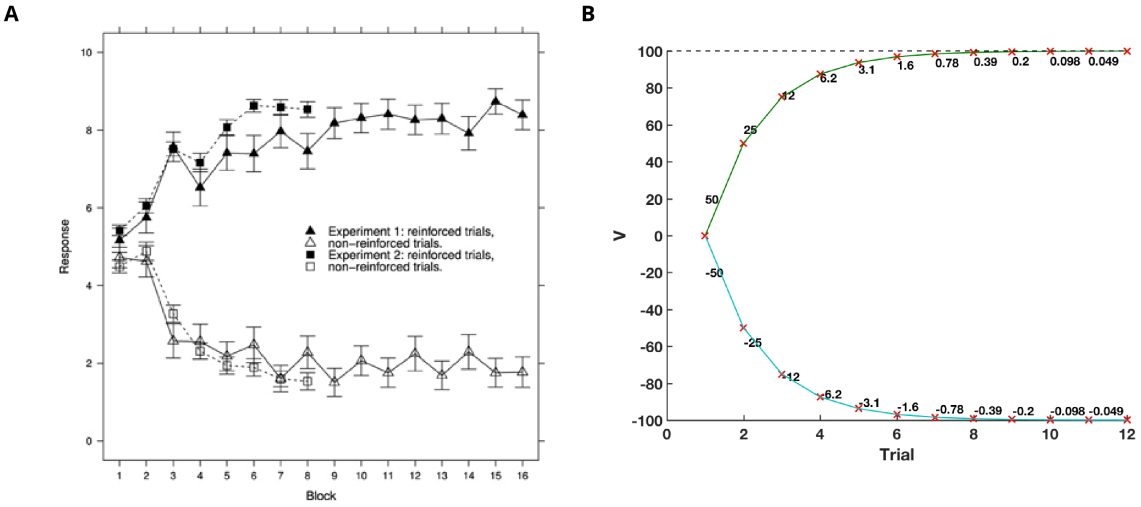


Figure 1.2 | Rescorla-Wagner captures classical conditioning (A) Acquisition curves for a task in which participants learned whether a card is rewarded or not. Blocks of trials are plotted against participants' ratings of reward probability. Adapted from [48] **(B)** The Rescorla-Wagner model produces decelerating learning curves, as observed in behaviour. Numbers correspond to learning rate * prediction error ($\Delta \boldsymbol{V_A}$) on each trial – see equation 1.3.

The RW model captures the intuition that we alter our beliefs in response to surprising events. It also explains several well-established phenomena. The first is Kamin blocking [49] in which the presence of established conditioned stimuli that perfectly predict a given US block the conditioning of new CS's. This is accounted for by the fact that the update term in equation 1.3 includes the prediction not only from the stimulus in question, but *all stimuli* in the environment; if another stimulus predicts the reward, the prediction error will be zero, and new associations will not be formed. The RW model also explains over-expectation [50], in which the compound presentation of two highly rewarded CS's (generating a very high expectation of reward) followed by the provision of a single dose of US leads to a decrease in the value of each CS (because the received reward is less than predicted), an effect found to depend upon the orbitofrontal cortex [51].

The curves produced by the RW model elegantly capture the acquisition rates displayed by animals learning CS-US relationships (Figure 1.2). The Rescorla-Wagner rule has thus provided a capable and influential account of how animals build a primitive model of the world, associating stimuli with their predictors.

1.3.1.2 Temporal-Difference Learning

Although the RW model enjoyed great success explaining trial-by-trial learning, it failed to explain learning on a more granular timescale. Richard Sutton and Andrew Barto noted that the RW-model fails to emphasise a feature of learning that had previously been stressed in engineering algorithms, that of prediction [52]. Both animals and artificial agents, they argued, are seeking to find the earliest, non-redundant predictor of a variable in which they are interested in, namely reward. Their account, Temporal Difference (TD) learning, extends the idea of prediction-error learning to continuous time, providing fine-grained predictions of within-trial dynamics to which RW-models are blind. Whilst RW-models describe the association between a predictor and an outcome, TD-learning seeks to find the earliest possible predictor of reward. This captured essential temporal components of learning which the RW-model was unable to account for, such as the importance of within-trial timings for CS-US associations [52]. The TD model also introduced the notion of a stimulus trace, which allowed predictive signals that occurred some time ago to be linked with the occurrence of reward at a later date.

They envisaged that an agent's predictions of future reward are encapsulated in a state value, which captures all of the (time-discounted) rewards that the agent expects to receive in subsequent states:

$$V_t = \sum_{i=1}^{\infty} \gamma^{i-1} r_{t+i}$$

Equation 1.4

Here V_t is the value of the current state. It is obtained by summing the product of reward on each timestep (r_{t+i}), discounted by the factor γ. γ is constrained to be less than 1 and raised to the power of the future timestep, i, thus producing an exponentially decaying influence of future rewards.

The core insight is that this quantity, V_t, can be learned in an incremental fashion by updating the previous time-step's prediction of reward to match the next time-step's, whilst taking into account any reward received on that timestep:

$$V_t \leftarrow V_t + \alpha(V_{t+1} + r_{t+1} - V_t)$$

Equation 1.5

Where V_t is updated to reflect knowledge about state V_{t+1}. As before, α is a learning rate parameter that determines how much the state value is updated on each timestep, and r_{t+1} is immediate reward. Note that, like the Rescorla-Wagner model, TD-learning relies upon a prediction error to drive learning:

$$\delta_t = V_{t+1} + r_{t+1} - V_t$$

Equation 1.6

The only difference with the RW prediction error being the presence of the term V_{t+1}, which reflects the notion that value updates should reflect not only the reward received by transitioning into a given state, but also all of the future reward associated with the new state. For instance, if you are the second son of the Tudor King of England, the death of your older brother may be associated with some small reward – his horse, and perhaps his wife. Far more significant is the value of the state in which you now find yourself, that of heir to the throne, associated as it is with innumerable opportunities to extract taxes, host lavish banquets, and wage war upon the French.

1.3.1.3 Reward Prediction Errors in the dopamine system

This term, δ_t, is known as the Reward Prediction Error (RPE), and has come to occupy a central role in our understanding of learning and the brain. Fundamental to the widespread adoption of the RPE by the neuroscience community was the revelation in 1997 that dopamine neurons in the Ventral Tegmental Area (VTA) and Substantia Nigra pars Compacta (SNc) appeared to represent precisely this quantity [53]. Neurons containing dopamine are densely concentrated in the SN/VTA, receiving inputs from a wide variety of structures [54] and projecting extensively to both cortical and subcortical structures, notably including the striatum and large portions of the prefrontal cortex (PFC) [55]. They thus appear well-positioned to broadcast prediction error signals to brain regions known to be important for representing and updating value estimates [26].

In probably the most frequently reproduced panel in computational neuroscience, Schultz and colleagues asserted that dopamine neurons align their responses to the earliest possible predictor of reward, and incorporate a temporal prediction such that the failure to receive a reward at the expected timepoint is associated with a reduction in firing (Figure 1.3A). The question posed in that panel – 'do dopamine neurons report an error in the prediction of reward?' – has been answered in the affirmative by two decades of work in rodents [56,57], monkeys [58-62], and humans [63-67]. Notable empirical successes include the observation that dopamine neurons are sensitive to blocking [59], respond to cues predicting reward omission [62], and weight previous rewards in an exponentially decaying manner [58], all as predicted from theory. The recent advent of optogenetic techniques [68] has allowed direct manipulation of neurons, confirming that synthetic prediction errors substitute for positive [56] and negative [69] RPEs (Figure 1.3B). Delicate dissection of the local circuitry within the SN/VTA suggests that inhibitory neurons containing the neurotransmitter GABA provide an expectation signal to dopamine neurons [57], which can be exogenously manipulated to produce dopaminergic and behavioural signatures of reward expectation [70] (Figure 1.3C). It should be stressed, however, that TD-learning does not provide a comprehensive description of the responses of dopaminergic cells, which display considerable variability in their responses [71] according to their inputs [72] and outputs [73], and the reward context in which they are assessed [74]. Further evidence suggests that some behaviours that might naturally have been subserved by reinforcement learning do not depend upon dopamine at all [75,76], whilst manipulating dopamine in the absence of rewarding stimuli does not seem to produce learning [77]. A full discussion of the range of responses of dopamine neurons is beyond the range of this thesis – for recent reviews see [78,79].

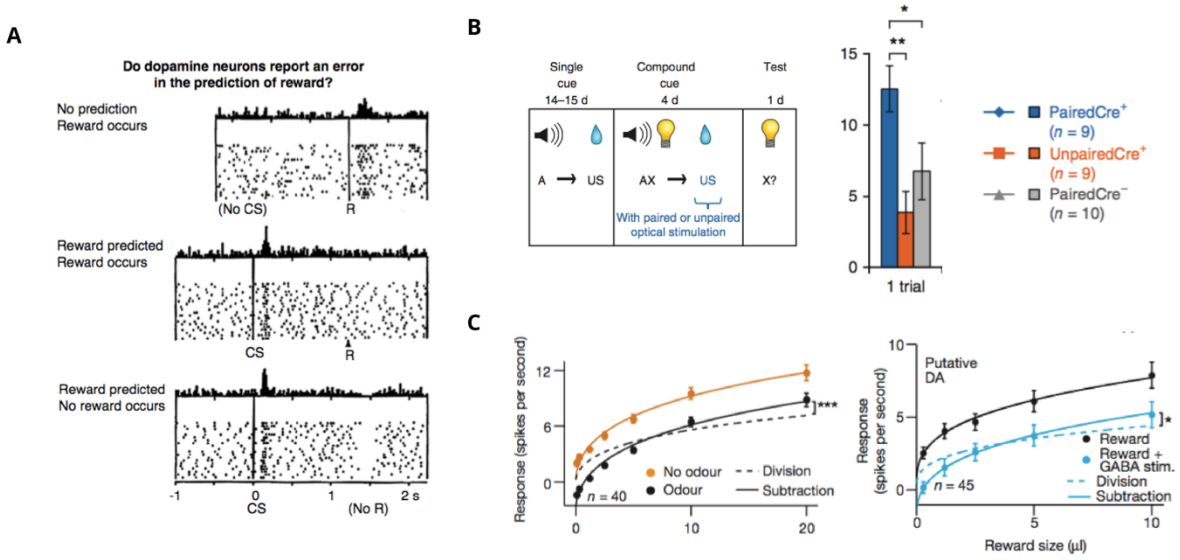


Figure 1.3 | Dopamine neurons convey an RPE signal (A) Dopamine neurons fire to the presentation of an unexpected reward (R) (top). When a CS is associated with a reward, dopamine neurons shift their responses to the CS, displaying no modulation of activity at the receipt of reward (middle). Omission of reward leads to a dip in the firing rate of dopamine neurons, consistent with a negative prediction error (bottom). Adapted from Schultz et al [53] **(B)** Causal evidence that dopamine neurons provide a signal that supports learning. Steinberg et al trained mice on a tone-juice pairing. This produced blocking, such that training with a compound stimulus (tone + light) did not produce learning of a light-juice association. Blocking was alleviated by pairing presentation of the US with optogenetic activation of dopamine neurons in the VTA (blue bar). Stimulation of dopamine neurons during the ITI did not produce this effect (red bar), demonstrating the temporal specificity of the learning signal. Grey bar is a control condition with a non-functional optogenetic vector. Adapted from [56] **(C)** Expectation exerts a subtractive influence upon dopaminergic responses to reward delivery, consistent with the formulation of the RPE in TD-learning. Eshel et al further demonstrated that local GABAergic cells appear to represent this expectation, providing subtractive inhibition to dopamine cells in the VTA. Adapted from [70].

1.3.2 From prediction to action

The responses that Pavlov observed in his dogs were involuntary. As we have seen, TD-learning provides a description of how the physiological responses associated with food– such as salivation – might be realigned to a predictive cue. This says nothing, however, about what the dog might be able to do to *obtain* food. Outside of the laboratory, passively waiting for the delivery of food is not a good strategy. The brain thus needs a mechanism for learning what to

do, not merely what to *expect.* The acquisition of behaviours that lead to rewarding stimuli and avoid punishing ones is referred to as *instrumental conditioning*.

Thorndike conducted a series of experiments in which animals had to solve problems, such as finding their way out of a puzzle box, to obtain a reward [80]. His work, and subsequently that of Skinner, established that animals were able to associate actions with their effects, thus making them more or less likely to repeat those actions in the future. Skinner showed that animals (such as his star laboratory rat, Pliny) could be coaxed to perform elaborate sequences of actions through repeatedly rewarding component actions, a process described as *shaping* [81]. He further coined a term to describe how outcomes shaped actions: reinforcement.

Reinforcement learning describes algorithms by which organisms can select actions to maximise a numerically defined reward[6]. They have enjoyed widespread use in neuroscience and computer science, where novel instantiations continue to extend the power of artificial agents to learn and act in novel environments [11]. The TD learning algorithm discussed above provides a reinforcement learning algorithm for the *estimation* of state value, and can be easily extended to describe how such estimates can guide action *selection*. Here we discuss one such instantiation, Q-learning, and summarise the evidence that the brain uses reinforcement learning algorithms to guide action selection.

1.3.2.1 Q-Learning

Q-learning builds upon the principles of TD-learning to describe how an organism can learn the value of different actions through trial and error [82]. Q-learning allows an agent to derive a *policy*; a way of acting in a set of states. It deals with a particular class of problems known as Markovian Decision Processes (MDP). MDPs possess the Markov property: the probability distribution over future states depends solely upon the current state. One doesn't need to remember what actions you took in the past or which states you've visited. This means that an agent can summarize how good a state is – and how good a given action in a given state is – with a single number. Luckily, many situations with which an organism has to contend enjoy the Markov property. Clever representation of the state space can also render non-Markovian decisions Markovian, by explicit representation of partially observable variables, a process that depends upon the OFC in humans [83].

Q-learning gives us a way of determining our policy when we are ignorant of two key pieces of information: the environment's reward structure (*V*) and its transition structure (*P*), which determines how we move between states. Instead, we seek to maximise rewards by estimating the value (*Q*) associated with a given action and state pair and choosing accordingly.

Watkins and Dayan [82] showed that the agent can learn this Q-value, and thus determine the correct policy, by iteratively updating its estimates in a manner reminiscent of temporal difference learning. At each timestep (*t*), the agent finds itself in a state (s_t) and takes an action (a_t) according to its current policy. It then evaluates its best estimate of the new state's Q-values, and updates the Q-values of the previous step according to the highest of the new Q-values. Thus at time step t:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

Equation 1.7

Where as before, γ is a discount function and α is a learning rate. Happily, following this update rule eventually leads us to find the optimal Q values, often referred to as Q* [82]. Q-learning thus makes a remarkable promise: simply by following a straightforward update rule and acquiring enough experience, an agent can learn to perform optimally in an environment about which it initially knows nothing.

Convergent evidence from humans and other animals suggest that the striatum is the most likely substrate of this kind of action-value learning in the brain [84]. Early work established that lesions to the basal ganglia abolished the ability to associate stimuli with responses, analogous to a failure to learn Q-values [85]. Subsequent recording studies in animals documented a representation of action-values in striatal neurons [86-88], with a parallel body of work in humans describing BOLD correlates of action-values in the striatum [64,89]. Given the links between dopamine and RPEs reviewed above and the dense dopaminergic projections to the striatum [90], it is intuitive to equate the striatum, particularly its dorsal part [91], to an action-selection system which receives a dopaminergic instructive signal [92]. Precisely how reinforcement-learning algorithms such as Q-learning are implemented in the brain remains a topic of vigorous research, and one too large to tackle here – for a comprehensive review, see [93] and [94].

1.3.2.2 Using Q-learning to understand behaviour

In addition to explaining neural responses, Q-learning has provided a useful model for behaviour. By providing a mechanistic account of how animals learn, Q-learning allows us to understand alterations of learning in terms of the processes described in reinforcement learning algorithms – prediction, evaluation, and updates. Model-fitting to behaviour can be used to understand the impact of pharmacological manipulations [95], the impact of brain stimulation [96,97], and understanding learned helplessness in depression [98] .

One further insight from the application of Q-learning in behaviour is that estimates of state-value, such as those discussed in the section on temporal difference learning, appear to affect behaviour in a manner not predicted from a pure Q-learning account [99]. A growing body of work from Guitart-Masip and colleagues asserts that there is a privileged relationship between positive state values and action, and negative state values and inaction [100]. In support of this claim, human participants struggle to learn an association between reward and inaction, display a tendency to withdraw from aversive stimuli, and dopaminergic responses to reward in the striatum are potentiated by action [101,102]. This can be understood through an evolutionary lens: situations associated with potential rewards are statistically associated with action, whereas avoiding threat often necessitates freezing or withdrawal. In **Chapter 3** we use a task designed to tease out the relative roles of valence and action to understand changes in learning under stress.

1.3.2.3 From model-free to model-based

Q-learning allows the agent to derive a policy in the absence of knowledge about the reward structure or the transition structure of the environment. We referred to this form of learning as model-free, reflecting the idea that the agent has no concept of the relationship between different states – there is no internal model of the world. This blunt trial-and-error learning clearly provides an impoverished account of cognition. As early as the 1930s, it was argued that animals stored internal models of their environment, and could use these to navigate to reward if called upon to do so [103]. Crucially, these models seemed to be acquired in the absence of direct reinforcement. In the language of machine learning, we describe algorithms that rely upon some knowledge about the structure of the world as being *model-based*: they require the organism to build some internal representation of the task, such as the transition structure *P* to

which we alluded earlier. They can then use this representation to guide their reinforcement learning by, for instance, evaluating each state on the basis of its connections to other states [104].

Model-based systems also confer the ability to do *planning* – to envisage a desired state, and determine how to reach it. Tasks designed to probe the relative contribution of model-free and model-based algorithms suggest that signatures of both are simultaneously present in the human brain [105,106]. Further investigation has suggested that the balance between the two can be altered by systemic manipulations of dopamine [107] and disruption of the dorsolateral PFC [108]. Recent work also suggests that arbitration between model-based and model-free systems is uncertainty-based and involves frontopolar cortex [12,109].

The distinction between blind model-free and prospective model-based systems has also been established in experimental psychology. Many decades of work have delineated the difference between habitual and goal-directed systems [110]. Although this literature developed without specifying algorithms by which these properties came about, the theory suggested that in one case the association was between a stimulus and a response (habitual, or model-free), and in the other between a response and an outcome (goal-directed, or model-based). The litmus test for the distinction between the two is reinforcer devaluation[111]. An animal is trained to press a lever in order to obtain a desirable outcome, such as a food pellet. The outcome is then devalued, either by pairing it with an undesirable flavour (such as quinine or saline), or through feeding to satiety. The diagnostic test is whether the animal continues to press the lever in order to obtain the outcome. If the learning is habitual – the stimulus (lever) provokes the response (press)- then devaluation does not alter responding, as the stimulus itself is unaltered. Conversely, goal-directed responding should be abolished if the goal is rendered undesirable. Devaluation has therefore provided a simple and elegant way to probe the nature of learning underlying choice.

Careful behavioural and lesion studies identified that the contribution of habitual responding tends to grow throughout training [112,113], is dependent upon lateral but not medial striatum [114], and is potentiated by dopaminergic drugs [115]. Conversely, goal-directed learning in rodents depends upon the integrity of the OFC and dorsomedial striatum [116-118]. Importantly, interindividual variability in the propensity to acquire habits, as assessed via devaluation, correlates closely with estimates of model-based behaviour [119], supporting the assumption that

these two schemas describe the same distinction. In this thesis we will encounter decisions thought to be underpinned by values from a model-free system in which the value of actions is learned over many trials (**Chapter 3**), and a model-based system, in which information about quantity and quality are flexibly combined (**Chapter 5**).

1.3.2.4 How far can we go with RL?

The beauty of reinforcement learning methods is that they reduce a daunting problem to one in which the aim is typically to track a single value via simple update terms. This is both intuitively appealing and computationally feasible. However, starting from a position of informational efficiency, this is a wasteful approach [120]. A strong line of argument from probability theory suggests that the brain would be better off representing beliefs as probability distributions, rather than single values [121].

To illustrate why this might be a useful feature, consider a coin flip. Imagine that we have bet on a heads, such that V_{heads}=1 and V_{tails}=0. We happen to know that the probability of a heads is 0.5 – in this case we have a *model* of the world that allows us to precisely infer the probabilistic structure. We might equally have learned this via model-free RL-learning, such that $V_{preFlip}$=0.5. Upon the resolution of this coin flip, we experience an RPE: either positive (+0.5, if heads) or negative (-0.5, if tails). Thus even in situations where there is nothing left to learn, we experience prediction errors. As one might predict from the discussion above, tasks analogous to coin-flips elicit robust activation in regions of the brain rich in dopaminergic innervation when the outcome is revealed [65,122,123] [67].

In RL-models, therefore, we have no explicit representation of uncertainty. Each time the coin flip is revealed we receive a prediction error, but our representation of $V_{preFlip}$ is identical to that of a situation in which we receive a reward of 0.5 every time! Assuming that we start with the assumption that the coin is unbiased – that is, our initial $V_{preFlip}$=0.5 – this also implies that after a 1000 coin flips, we have learned precisely *nothing*. We are now very *sure* that the coin is unbiased, but RL does not give us the ability to represent this confidence.

From a neurophysiological perspective, this seems wasteful: the brain is being bathed in dopamine, promoting plasticity, even in situations in which there is nothing left to learn [124]. Although suggestions have been made as to how RL models might be modified to account for

this [125], the general consensus is that this is a limitation of this class of models [126]. To incorporate an explicit representation of uncertainty, we must turn to another tool: that of Bayesian statistics.

1.3.3 Summary and relevance

Animals use algorithms based upon surprise to learn the association between stimuli, actions, and rewards. The Rescorla-Wagner model provides a trial-by-trial account of how stimuli come to be associated with rewards, whilst temporal-difference learning extends this process to find the earliest possible predictor of reward. Reinforcement learning algorithms, such as Q-learning, describe how organisms can learn to select actions to obtain rewards. Learning systems can be supplemented by models of the environment, permitting flexible behaviour and planning. However, RL models don't permit a representation of uncertainty, which appears a substantial shortcoming.

In **Chapter 2** I present a prediction-error model for fluctuations in subjective happiness, and describe its ability to predict choices about how much money to donate to other players. In **Chapter 3** I use a variant of Q-learning to describe learning following acute stress, detailing action-specific deficits in learning.

1.4 Uncertainty in the environment, decision-making, and the brain

It is worth pausing to underline that the type of decisions with which we are concerned here are those conducted under conditions of *uncertainty*. This encompasses the vast majority of (interesting) choices that an agent makes. To appreciate the pervasiveness of uncertainty, we can turn to the small subset of situations in which the link between sensory information and appropriate action is of such high fidelity that we can think of it as being uncertainty-free. The actions we produce in such situations are called reflexes, and they are unusual in a variety of ways. Firstly, there is one to one mapping between the input (such as activation of stretch receptors in the patella) and the output (contraction of the quadriceps and straightening of the leg). Secondly, they are fantastically stable over time, to the extent where they are genetically encoded rather than learned. The relationship between input and output is so reliable that evolution has baked it into the nervous system. Finally, and as a result of their simplicity, their execution is achieved entirely without the use of the brain at all, instead relying upon the spinal

cord. This might suggest that the elaborate machinery of the brain is largely dedicated to dealing with the (vastly preponderant) other situations, in which uncertainty abounds.

1.4.1 Bayesian inference

Although the brain has to deal with copious uncertainty, it usually has several sources of evidence on which to draw. For instance, speech comprehension doesn't just rely upon auditory information; we draw upon visual information about the movement of the lips and tongue and a rich body of contextual cues suggesting the likely intentions of the speaker, not to mention extensive knowledge of the structure and semantics of language. Bayesian inference tells us that the statistically optimal approach when faced with multiple sources of noisy information is to integrate them with a weight inversely proportional to their uncertainty. This requires us to represent not only a point-estimate of our belief, but a probability distribution (O'Reilly et al., 2012). As we shall see in subsequent sections, there are a variety of suggestions as to how the nervous system might achieve this [127].

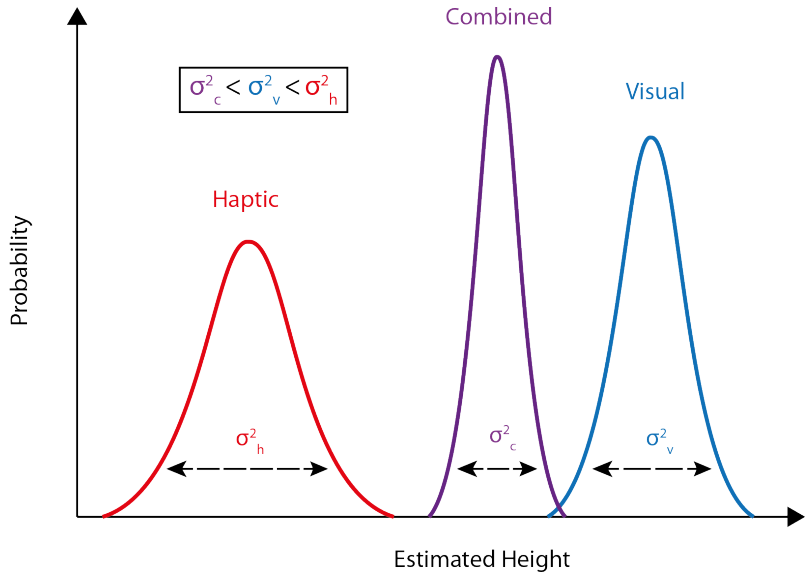


Figure 1.4 | Integrating multiple sources of evidence In Ernst & Banks (2002), individuals estimated the height of a bar according to noisy visual and haptic information. Integrating sources of evidence according to their uncertainty, here represented as the width of a probability density function (σ^2), produces a combined estimate with a lower variance than either of its constituents ($\sigma^2_{combined} < \sigma^2_{visual} < \sigma^2_{haptic}$). A model implementing Bayesian integration of this kind was shown to be a good predictor of individuals' performance on the task. Adapted from [128].

A Probability Density Function (PDF) describes the probability that a given variable will take on a certain value. Assuming for a moment that our probability distributions are Gaussian, the mean of the PDF is our best guess – the most probable value of the variable. The width, or variance (σ^2) of the distribution corresponds to the uncertainty associated with the representation of that variable (Figure 1.4). The combination of multiple sources of information produces a PDF in which the variance is smaller than that of either of the inputs (Figure 1.4), according to the equation:

$$\sigma_{total}^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Equation 1.8

Where σ_X^2 is the variance associated with the probability distribution *X*. This gives us a way of combining sources of evidence so as to arrive at the most accurate conclusions possible.

Whereas accounts of reward-learning have been dominated by approaches derived from reinforcement learning, Bayesian algorithms have been more popular in the field of sensorimotor control. There is convincing evidence that humans combine sources of information in this way, when estimating the height of a bar from noisy visual and haptic information [128] (Figure 1.4), the position of a noise source using visual and auditory cues [129], and in guiding movements [130]. This uncertainty-weighting of multiple information sources even seems to be applied when groups of individuals make decisions about a commonly observed perceptual event [131]. This suggests that accounts of reward-guided decision-making which omit a representation of uncertainty might be overly simplistic.

1.4.2 Bayesian learning

Bayesian inference is not merely useful for combining multiple sources of current sensory input. It also gives us a way to optimally incorporate our evidence with prior beliefs, formed as the result of learning. This is clear from the formulation of Bayes' theorem:

$$p(A|B) \propto p(B|A)p(A)$$

$$\textit{posterior} \propto \textit{likelihood x prior}$$

Equation 1.9

Or, to borrow a popular tennis example [132]:

$$p(ball\ at\ x|perception\ of\ ball\ at\ x) \propto p(perception\ of\ ball\ at\ x|ball\ at\ x)p(ball\ at\ x)$$

Our attempts at ball-localization are improved by taking into account not only the *likelihood* p(perception of ball at x|ball at x), which tells us how good our sensory evidence is, but also the *prior* distribution p(ball at x), which describes the typical distribution of tennis shots. The importance of having good priors is best illustrated in the return of serve. A returner has about 350ms to position their racket before an Ivo Karlovic serve becomes an Ivo Karlovic ace. In such extreme circumstances, high-quality prior information is an important determinant of success. The process of incorporating priors and likelihoods to form a posterior is illustrated in Figure 1.5.

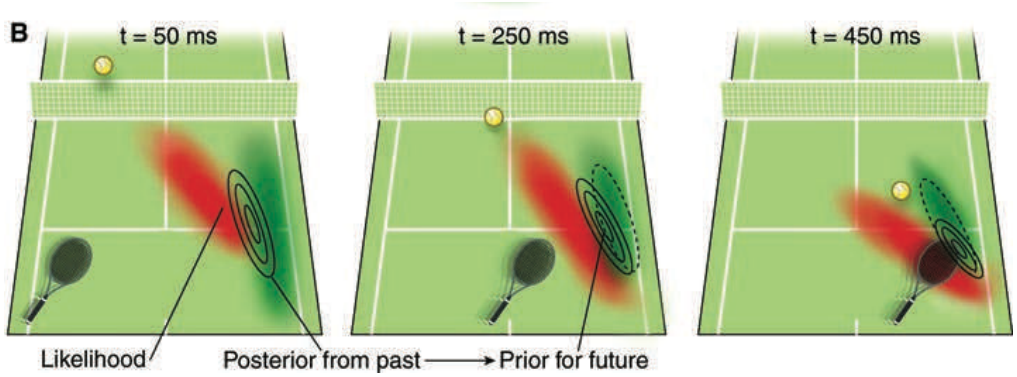


Figure 1.5 | Combining priors and likelihoods in space A Bayesian tennis player trying to predict the likely landing location of an opponent's shot can combine the observed trajectory with the prior distribution of shot placement. Adapted from [132].

Priors are established by learning; they are the result of experience. Bayes' theorem gives us an intuitive way to update our beliefs as a function of new evidence (the likelihood) and our existing beliefs (priors). Just as in sensory integration (see Figure 1.4), the two sources of information are represented probabilistically, and combined according the uncertainty associated with each. This is the mathematical equivalent of the truism 'extraordinary claims require extraordinary evidence'. Outlandish evidence that conflicts with our existing beliefs must be of very high quality in order to substantially influence our opinion.

Kording and Wolpert [130]provided persuasive evidence of this process in a sensorimotor learning task in which participants had to move their finger to a target. They were unable to see their

finger, but received visual feedback documenting its trajectory. Feedback was laterally displaced on each trial by an amount drawn from a Gaussian distribution, and the quality of feedback provided varied from low to high quality. Kording and Wolpert showed that people learned the prior distribution of possible displacements, and adjusted their reaches according to a combination of the prior and the feedback they received during the movement, weighted by the uncertainty associated with the latter.

1.4.3 Bayesian learning in volatile environments

The reliability of a prior belief may change over time at different rates. In dynamic environments, old beliefs should be rapidly overwhelmed by new evidence. In very stable ones, old information is still valuable. For instance, if you last followed football 10 years ago, your beliefs about which teams are good ought to be pliable in the face of this weekend's results. However, if you last lived in London 10 years ago, you can be reasonably confident that your predictions about the weather are likely to remain reasonably accurate. This is because of the difference in the *volatility* of the two scenarios: football is far more volatile than climate, meaning that the relationship between outcomes and their predictors is more likely to shift over time. A Bayesian learner is able to accommodate this by representing their beliefs as a probability distribution, with higher uncertainty reflected by a broader distribution over possible values. Behrens et al [47] showed that humans are able to track the volatility of an environment, allowing them to learn quickly in volatile situations, rapidly overwriting old beliefs, but relying upon longer reward histories when the environment was more stable. To frame their approach in the syntax of RL, they found that the learning rate – α in equation 1.8 - varied systematically with the volatility of the underlying statistical process, precisely as we would expect from a Bayesian learner. Indeed, comparing the ability of different models to account for participant behaviour demonstrated that an optimal Bayesian learner comprehensively outperformed an RL model, despite the latter being tuned to fit the data via free parameters. This Bayesian approach to learning has inspired several subsequent models, such as the Hierarchical Gaussian Filter [133] to which we will return in due course.

1.4.4 Delineating different forms of uncertainty

There have been several attempts to parse uncertainty into distinct forms[124,134]. The most intuitive distinction is between uncertainty that results from genuine unpredictability in the

environment and uncertainty that can be reduced by learning. The former is *irreducible uncertainty*: the organism cannot hope to reduce it by further learning. This is also commonly referred to as *risk* in psychology and behavioural economics. The unpredictability of a coin toss is a good example of irreducible uncertainty; you can study coin tosses for as long as you please, but your predictive accuracy is unlikely to depart from 50%[*]. Conversely, situations in which outcomes are governed by a rule that can be learned create *estimation uncertainty*, which is diminished by learning. This is a measure of ignorance about the predictive relationships in the environment. To return to the tennis example, learning that your opponent favours shots to your backhand reduces uncertainty about where the ball will land, via a reduction in estimation uncertainty. But what if your opponent realizes that you've noticed this predictability, and shifts his shot placement to favour your forehand? Such instability will necessitate relearning. This is an example of the *volatility* referred to earlier. Some have suggested that this injection of uncertainty is best described as *unexpected uncertainty*, in contrast with those forms of uncertainty which remain relatively stable over time [124]. An alternative is to regard volatility as merely another layer of probabilistic relationships, which instead of affecting behaviour directly, determine the likelihood that behaviour will change [133]1. From this perspective, we can invoke a third kind of uncertainty, which is the uncertainty over the volatility in the environment; *volatility uncertainty.* In fact this is a form of estimation uncertainty, the only difference being that the value being estimated is not the probability itself, but its instability.

These multiple forms of uncertainty rarely exist in isolation. One of the key challenges faced by a learner in a dynamic environment is the simultaneous estimation of each kind of uncertainty, the balance between which is crucial in assimilating experiences to guide future decisions. It is easy to acknowledge that things are unpredictable: deciding whether that is due to genuine stochasticity (irreducible uncertainty), lack of learning (estimation uncertainty), or because things are rapidly changing (volatility uncertainty) constitutes a major challenge. One successful theoretical approach is to organize the different quantities to be tracked in a hierarchy, with

[*] The astute (pedantic) reader might point out that, from the perspective of Newtonian mechanics, the coin is in fact predictable. Here we use irreducible uncertainty to describe information that is permanently inaccessible to the agent; this uncertainty is, to all intents and purposes, irreducible.

estimates at each level associated with their relative uncertainties [133]. The uncertainty representations in this model, the Hierarchical Gaussian Filter, naturally recapitulate the notions of irreducible, estimation, and volatility uncertainty described above.

1.4.5 The Hierarchical Gaussian Filter (HGF)

The HGF is, like the Rescorla-Wagner, a predictive model. In the version we consider, the aim is to predict a binary outcome (u) on the basis of being in a particular state. Practically, we will consider the case where there are two states (such as hearing one of two tones), and we are trying to predict which outcome (such as a picture of a house or a face) is likely to follow, given a fluctuating relationship between states and outcomes.

Unlike the RW-model, the HGF poses the learning problem as a problem of function approximation, and endows the learner with some beliefs about the likely forms of the approximated function. The name of the model comes from the form of these functions: they are a series of hierarchically arranged Gaussians, with the learning rate of Gaussians lower in the hierarchy determined by the value of Gaussians higher in the hierarchy. Here we sketch out the structure of the model and its previous implementations: for an account of its implementation and interpretations of each parameter, see **Chapter 4**.

In the instantiation we consider, the model consists of three levels, each corresponding to a probability distribution (Figure 1.6A). The top two are Gaussians, whilst the bottom level is a binary prediction, derived from a sigmoid transformation of the second level. The ultimate aim of the learner is to estimate the sufficient statistics of these distributions, namely their means and variance (Figure 1.6B). This is achieved via a variational approximation, resulting in an update scheme for each level that bears a striking resemblance to the prediction error updating seen in RL. Importantly, the learning rate on these updates is allowed to vary as a function of uncertainty, leveraging the key advantage of Bayesian models over their RL counterparts (1.6C).

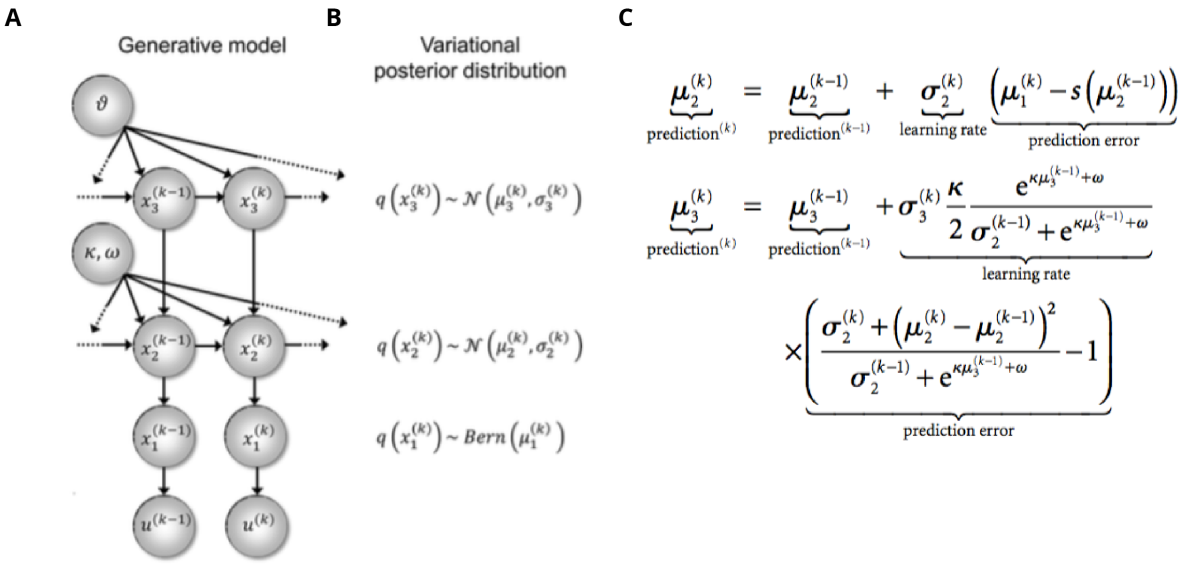


Figure 1.6 | The Hierarchical Gaussian Filter (A) Left: The generative model hypothesized to underlie observations (u). The hidden layers $X_{2\text{-}3}$ are Gaussians, whilst X_1 is a Bernoulli distribution defined by the value of X_2. In the generative model, an observation results from X_1 on each trial. **(B)** Sufficient statistics parameterising each layer. The agent tracks the means ($\boldsymbol{\mu}$) and variances ($\boldsymbol{\sigma}$) associated with the distributions in each layer. This is achieved by updates via prediction errors. **(C)** Update terms illustrating how upwards-propagating prediction errors allow learning at each level. Note that the learning rate at each layer is a function of uncertainty about the current value of the distribution ($\boldsymbol{\sigma}$). Adapted from [133].

In order to appreciate how the HGF incorporates uncertainty into the update equations, let us examine in detail how updates are performed. On each timestep, the model produces a prediction at each level, distinguished by the presence of a circumflex (^). This produces the predicted outcome on that trial:

$$\hat{\mu}_1^{(k)} = s(\mu_2^{(k-1)})$$

Equation 1.10

Where k is the current trial and s is the sigmoid transform. It is at this point that the uncertainties at each layer incorporate the value of the layer above:

$$\hat{\sigma}_1^{(k)} = \hat{\mu}_1^{(k-1)}(1 - \hat{\mu}_1^{(k-1)})$$

Equation 1.11

$$\hat{\sigma}_2^{(k)} = \sigma_2^{(k-1)} + e^{\kappa\hat{\mu}_3^{(k-1)}+\omega}$$

Equation 1.12

$$\hat{\sigma}_3^{(k)} = \sigma_3^{(k-1)} + \vartheta$$

Equation 1.13

Where κ, ω and ϑ are all subject-specific parameters. κ determines the degree of linkage between second and third layers. ω is a constant contribution to the learning rate at the second level. ϑ is a fixed estimate for a meta-volatility parameter, capturing beliefs about how frequently volatility changes occur. As we shall see, ϑ is affected by chronic stress (**Chapter 4**).

We are now in a position to appreciate how the HGF relates to the RL models we encountered earlier. The second level of the model represents the probability distribution governing the state → outcome relationship. The mean of this distribution (μ_2) reflects the current estimate of that relationship, and is thus analogous to the expectation term in an RL model. However, whilst the learning rate in RL models is fixed, the learning rate of the HGF is malleable, as in the model of Behrens et al [47]. Recall from Figure 1.6C that updates occur according to the equation:

$$\mu_2^{(k)} = \mu_2^{(k-1)} + \sigma_2^{(k)}(\mu_1^{(k)} - s(\mu_2^{(k-1)}))$$

Equation 1.14

Where $\mu_1^{(k)} - s(\mu_2^{(k-1)})$ is the difference between the outcome and the prediction (i.e. the prediction error) and $\sigma_2^{(k)}$ is the learning rate, which is itself updated as:

$$\sigma_2^{(k)} = \sigma_2^{(k-1)} + e^{\kappa\hat{\mu}_3^{(k-1)}+\omega} + 1/\hat{\sigma}_1^{(k)}$$

Equation 1.15

The current learning rate thus reflects the uncertainty in our existing estimate $\sigma_2^{(k-1)}$, the volatility of the environment ($e^{\kappa\hat{\mu}_3^{(k-1)}+\omega}$), and the expected size of the prediction error from the lower level $1/\hat{\sigma}_1^{(k)}$, such that prediction errors are downweighted if we expect them to be large [125]. This means that we learn more from a large prediction error which is unlikely given our current beliefs, whilst learning dwindles in scenarios in which the irreducible uncertainty is assessed to be high, such as the coin flip discussed earlier.

The HGF has been successfully applied in both statistical [135] and social [136] learning tasks. In **Chapter 4** we will use the HGF to model learning in an aversive learning task. We will then test the impact of different forms of uncertainty, as estimated in the HGF, upon the evolution of stress responses.

1.4.6 Representing uncertainty in the brain

Although the behavioural evidence reviewed in section 1.4.1 strongly suggests that organisms use Bayesian-esque algorithms in learning and decision-making [137], the details of how the brain represents probability distributions are still unknown [127]. We can broadly classify models of probabilistic representations into two kinds: summary statistic or population coding accounts [138] (Figure 1.7).

Work in the neuroeconomic tradition has documented numerous correlates of the *summary statistics* of distributions. The central idea is that most probability distributions have a known form – such as Gaussian, or Poisson – and can be parameterized by their sufficient statistics. By representing these summary statistics directly, we avoid the costly process of representing the full probability distribution. Most work has focused on the representation of the first two moments of reward distributions: the mean and the variance. Both MRI and single-neuron studies have provided evidence that risk is explicitly represented in frontal and parietal cortices during choice [139-145], in concert with coding in the basal forebrain [146-148]. However, disambiguating signals relating to utility from those relating to risk has proved problematic, as risk itself affects utility (see section 1.2.1) [149,150]. For instance, McCoy & Platt [142] report that neurons in the posterior cingulate are sensitive to risk. However, in their experiment value was held constant and monkeys systematically preferred the risky option. It is thus unclear whether neurons reflected the characteristics of the reward distributions or the subjective preferences of the animals, since both risk and preference were higher for the risky options.

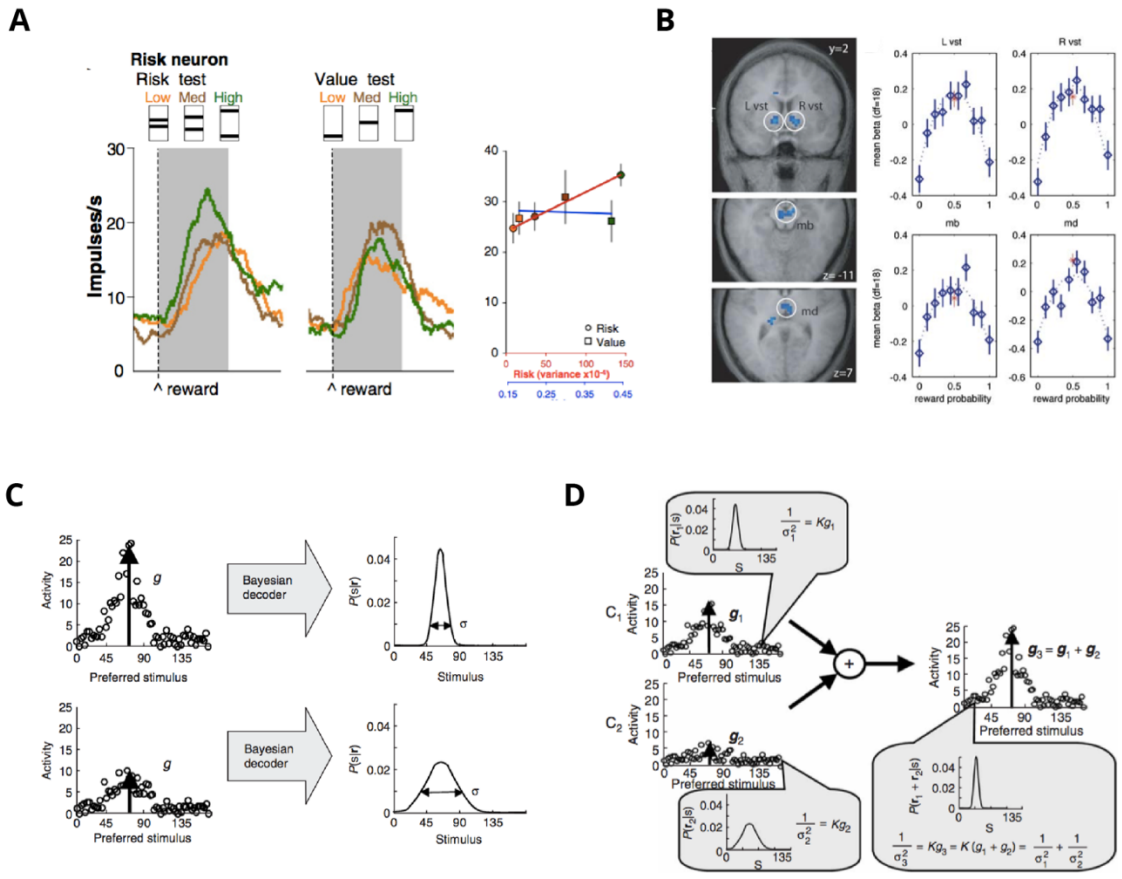


Figure 1.7 | Theoretical and experimental work on uncertainty coding (A) Representation of reward uncertainty (risk) in single neurons. Single-neuron studies frequently identify single neurons whose activity linearly encodes uncertainty (left panel). These neurons are usually distinct from those encoding reward itself (right panel). Adapted from [139] **(B)** Human neuroimaging studies also report correlates of risk in structures involved in choice. Here activity in the ventral striatum, midbrain, and mediodorsal thalamus is plotted as a function of reward probability; recall that uncertainty is highest when p=0.5. Adapted from [140]. **(C)** Probabilistic population codes. Ma and colleagues[163] proposed a model in which populations of neurons represent probability distributions over some variable of interest. Each neuron has a Gaussian tuning curve, and neurons are arranged by their preferred stimulus i.e. the Gaussian's mean. The gain of population response (*g*) represents the inverse of uncertainty, such that higher gain is decoded as slimmer posterior probability distributions (top row). Changing the gain of the population response corresponds to changing the variance on the resulting probability distribution (bottom row). Adapted from [163]. **(D)** Bayesian inference with probabilistic population codes. Here two population codes (top and bottom) are combined by simple linear summation. The resultant code (right) has a gain which is the sum of the two component codes, leading to a posterior distribution with the appropriate width, as calculated using Bayes rule. Summing population codes thus performs optimal Bayesian inference. Adapted from[163].

Single-neuron studies that identify *separate* populations of cells sensitive to value and risk in the OFC [139], and those that attempt to dissociate uncertainty signals from expected value [144], provide the strongest evidence for discrete summary statistic coding. However, subsequent work has called into question whether neurons in the OFC truly represent uncertainty, as opposed to salience. Ogawa et al [151] found that the vast majority of uncertainty-sensitive neurons also displayed differential responses to cues predicting the deterministic presence or absence of reward. They interpreted this in light of attentional models of learning [152] suggesting that such neurons are instead sensitive to salience, which is higher for both unpredictable and valuable stimuli [153].

Neuromodulatory representations of uncertainty have also been widely posited [124,154], although there is something of a ravine between the popularity of these ideas and their empirical support. Indirect evidence from measures of pupil diameter suggest a representation of uncertainty in the noradrenergic system [155-157], although direct confirmation is lacking. It also seems reasonable to suppose that diffuse modulatory systems are better suited to broadcasting the result of some computation, such as the estimated uncertainty, rather than being intimately involved in that computation per se [158]. Such neuromodulatory signals have also been posited on the basis of spiking neural network models of Bayesian learning [159].*

A second account emphasises how *populations*, rather than single neurons, might represent probability distributions as opposed to point estimates [162]. In the most influential of such models, Ma et al [163] forged a link between the kind of variability observed in neurons, Poisson noise, and the representation of uncertainty. The key insight is that since neurons are noisy – they respond to the same stimulus with differing firing rates- we can quantify the response (r) to a stimulus (s) as a probability distribution: $p(r|s)$. We know from Bayes theorem that this can be massaged into an estimate of the probability distribution for the stimulus, given a response:

$$p(s|r) \propto p(r|s)p(s)$$

* Some deep learning models adopt a summary statistics scheme – encoding the mean and variance of a Gaussian or a Bernoulli distribution – but do so using a representation distributed over many neurons [160,161]. To my knowledge, this idea is yet to receive attention in accounts of probabilistic representation in the brain.

Equation 1.16

The Poisson distribution is defined as:

$$p(k\ events\ in\ interval) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Equation 1.17

Where λ is the average number of events per interval, often referred to as the rate parameter. This can be thought of as the average firing rate evoked by a given stimulus, $f_i(s)$. Substituting in number of spikes (r_i) for k and firing rate ($f_i(s)$) for λ, we arrive at a precise specification of the term p(r|s):

$$p(r|s) = \prod_i \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!}$$

Equation 1.18

Omitting a normalization term for clarity, we now have a full description of the posterior distribution, $p(s|r)$:

$$p(s|r) = \prod_i \frac{e^{-f_i(s)} f_i(s)^{r_i}}{r_i!} p(s)$$

Equation 1.19

Which defines a population code that recovers Gaussian distributions for $p(s|r)$, with the peak aligned to the most probable stimulus value and variance inversely proportional to the gain of the constituent responses. These codes have the pleasing property that their linear addition constitutes optimal Bayesian inference, producing posterior codes that reflect the contributions of their constituents to a degree inversely proportional to their uncertainty. However, this code can only use gain to communicate uncertainty if neurons are gain invariant – meaning that their tuning curves do not change form if the gain of the response changes [163]. If the average population response varies over the range of stimulus values then the mechanism breaks down, because changes in stimulus value become indistinguishable from changes in stimulus uncertainty. This suggests that linear codes, such as those typically associated with value coding [40], are unable to support probabilistic population coding as envisaged by Ma and others [127]. In accordance with this observation, evidence for probabilistic population coding is strongest for

motion coding in MT [164] and representation of orientation in V1 [165], both of which are known to rely upon Gaussian tuning.

A second flavour of population coding suggests that variability in population activity *over time* represents uncertainty [137]. At any timestep, the population encodes a point estimate of the value, but temporal variability in firing rates encodes uncertainty. On a single neuron level, this is achieved if neurons are drawing samples of their firing rate from the posterior distribution they seek to represent [137]. This version of population coding is therefore known as the *sampling* account. It is clearly distinct from the summary statistic and probabilistic population coding discussed above, because it requires information about uncertainty to be represented by patterns over *time*, rather than providing an instantaneous readout by patterns of information over *space* (separate neurons). This has made it considerably more challenging to test, particularly in humans. Indirect evidence for sampling come from activity in the ferret visual cortex [166], in which the spontaneous distribution of activity over neurons comes to resemble the stimulus-evoked activity over neurons, which the authors suggest reflects the learning of a prior over stimulus distributions. Other groups have argued, however, that such patterns emerge as a consequence of global population dynamics, rather than reflecting the specific structure of activity over neurons [167]. Recent work has further developed the model in terms of visual cortex, providing predictions of the mean and variance of firing rates that accord well with data in V1 [168]. However, such accounts depend upon the fast dynamics – and thus low autocorrelations – of neurons in primary sensory areas [168], suggesting that they are inappropriate for the description of activity in the brain areas in which we are primarily interested, such as the prefrontal cortex, which display much slower dynamics [169] that have clear relationships to their role in cognitive processes [170].

1.4.7 Summary and relevance

Uncertainty is an inescapable feature of an agent's interaction with its environment, affecting perception, decision, and action. Bayesian statistics provides an elegant way to deal with uncertainty, and humans appear to rely upon Bayesian inference for sensorimotor learning and decision-making, combining information according to its uncertainty. Recent models have extended this idea to value learning, showing how the volatility in a value source can alter the speed at which people perform action-value learning. This has led to hierarchical models of

learning, such as the HGF, in which multiple forms of uncertainty – such as uncertainty stemming from risk, ignorance, and volatility- are separately represented and updated in a prediction-error framework. In **Chapter 4** we use the HGF to explain fluctuations in physiological and subjective stress responses during aversive learning.

Summary statistic approaches have dominated our understanding of the representation of probability distributions over value. However, evidence for such representations is not without issue, with uncertainty frequently confounded by expected value and salience. Furthermore, summary statistic approaches yield little understanding as to how the brain might perform uncertainty-modulated value learning, or incorporate uncertainty estimates into choice. Conversely, probabilistic population codes provide a more mechanistic account of computation under uncertainty, but have received little attention in value-based decision making. Existing knowledge about neural coding of value suggests linear coding schemes that are incompatible with probabilistic coding schemes, which have provided compelling characterizations of uncertain sensory processing. However, absence of evidence for non-linear codes is not evidence of absence; as we shall see in **Chapters 5 and 6,** careful dissection of value signals with methods sensitive to non-linearity suggest that value representations in the ACC and OFC might in fact be compatible with probabilistic population coding.

1.5 Emotions: causes and consequences

Much of the research discussed so far draws upon ideas from economics and machine learning, prescribing ways in which people ought to learn and behave according to normative accounts. An additional feature of animal decision-making is that it is sensitive to [*], and produces fluctuations, in emotion [172].

Although early debate raged about the relative contributions of psychological and physiological factors to emotion [173-175], contemporary efforts have focused upon the characterization of emotion's causes and effects. In particular, the contribution of cognitive science has been to describe emotion as a consequence of other cognitive processes, such as the recognition of

[*] Indeed, normative methods such as reinforcement learning provide a better fit to behaviour when estimates of emotional responses are incorporated [171]

threat or the receipt of reward [176]. It is this scheme which we adopt in this thesis, setting aside the various definitional issues that still plague the field, such as whether emotions are neatly divisible into a series of unique states, as argued by William James and later Paul Ekman [177], or whether emotional state is better described as a point in a continuous space defined by the axes of valence and arousal [178]. We first sketch out an influential theory describing why emotions are best thought of as feedback mechanisms, and discuss evidence that emotions are linked to the internal variables we have posited in our discussion of learning and decision-making. We then review the evidence that emotion can affect learning, and see that one way in which this occurs is through shifting the balance between different learning systems.

1.5.1 Emotion as feedback

Baumeister and colleagues [179] reviewed evidence that consciously experienced emotions chiefly function to modify future behaviour as a consequence of experience. Although early thinking in the field suggested that emotions triggered behaviour directly, Baumeister and colleagues argue that the florid emotional experiences with which we are consciously familiar occur too slowly to be a useful cause of anything. Relying upon the sensation of fear to initiate escape from a bear would be a sluggish strategy. Furthermore, the argument deployed against James' original theory – that the range of emotional states is too broad to be conveyed by a limited physiological repertoire- can be employed in reverse here, noting that the range of emotional states is *not wide enough* to provide a useful specification of action. Conversely, emotions develop on a timescale more suited to the modification of future behaviour by recent experiences. They might do this by biasing learning processes, such as improving the consolidation of recently formed memories [180,181], or by providing explicit memories of subjective states (e.g. guilt) that people then seek to pursue or avoid.

Evidence from the latter comes from ingenious 'mood-freezing' experiments, in which participants take a sugar-pill which they are informed will render them impervious to mood changes for the next few hours. This reveals behaviour that is prospectively geared to elicit mood-changes, such as helping behaviour that seeks to alleviate sadness [182]. Phrased in the language of previous sections, we can see that some of the effects of emotion act via model-free mechanisms (implicit alterations to reinforcement learning processes), and others through

model-based mechanisms (explicit simulation of future emotional state)*. The latter has received a good deal of attention under the banner of 'hedonic forecasting', the study of how people anticipate their future emotional state [183]. Such 'premotion' is vulnerable to systematic biases, many of them paralleling those observed in everyday decision-making, such as a bias towards salient events [184].

In **Chapter 2** we will see how participants faced with a novel social decision act in a way that reflects their emotional preferences, as inferred from a separate task. Although the mechanistic link between these two processes remains to be elucidated, this demonstrates that inter-individual differences in preferences are linked to inter-individual differences in emotional processing. Importantly, previous attempts to explain such social decisions without recourse to emotion have failed [185], suggesting that understanding individual emotional profiles is key to understanding social decision-making. Conversely, in **Chapter 3** we will document an effect of stress upon reinforcement learning, illustrating a model-free way in which emotion can impact behaviour.

1.5.2 Models of emotion

Having acknowledged that emotions influence both ongoing learning and future behaviour, we arrive at an important follow-up question: how is current emotional state determined? Although it seems obvious that fear is an appropriate reaction when confronted by a bear, what mechanisms specify how I feel when I open my bank statement, or read the news, or go to the pub? Carver and Scheier proposed that emotions gave a continuous readout of the state of the sort of goal-attainment processes we encountered in earlier sections [186]. Rapid progress towards a goal is reflected by positive emotions, whilst progress that is slower than expected produces negative emotions [187]. Oatley and Johnson-Laird expressed a similar idea, framed in terms of the likelihood of attaining a goal. Increases in this likelihood lead to positive, and decreases to negative, emotions [188]. In fact, the idea that emotion might reflect the status of an internal

* This is isomorphic to devaluation procedures discussed in the context of model-based control. In both cases, the outcome is rendered unattractive, and the experimenter asks whether behaviour putatively motivated by the outcome persists.

learning system can be traced back to Masanao Toda's eccentric thought experiments concerning the design of a fungus-eater on the planet Taros [189].

It is clear that this early work [186,188,189] described the generation of emotions in the same language that we have already encountered, that of goals, expectations, and error signals. However, without the organising framework of reinforcement learning, quantitative specification of these terms, including the question of how an organism's goals are defined, was lacking. RL provides a computational framework in which such ideas can be cleanly formalised: goals are specified by the value function, and positive and negative prediction errors reflected circumstances that are better or worse than expected. Rutledge et al [122] produced the first quantitative model of subjective well-being, demonstrating that positive and negative prediction errors in a gambling task produced reliable increases and decreases in positive affect in both the lab and in smartphone-based experiments. Neuroimaging confirmed that prediction error signals in the striatum were predictive of changes in subjective well-being. Insula activity at the time of measurement correlated with current well-being, echoing the well-established role of the insula in interoceptive and emotional processing [190,191].

Learning models therefore allow us to quantify hidden variables thought to be important for value-learning, and test their putative relationship to emotion. They also allow more detailed theorizing on the function of emotion, by asking what computational role they might serve in the learning processes from which they result. By characterizing the features of emotional processes, we can ask how they relate to the statistics of the environment. For instance, one characteristic of emotions is that they tend to change relatively slowly. The influence of past events upon subjective well-being decays exponentially, reflecting prediction errors over the last 4-5 trials [122]. This suggests that emotions might be integrating information over a longer timescale than envisaged by RL-type models.

1.5.3 Function of emotion

Eldar and colleagues proposed that fluctuations in emotion, or mood, provided a 'momentum' signal. They link the observations above – that mood changes relate to positive or negative prediction errors - with the finding that mood tends to bias assessment of value, with positive mood associated with overvaluation and negative mood with undervaluation [192-194]. They

suggest that this bias reflects the structure of real environments, in which changes in reward value are typically correlated. For instance, changes in rainfall produce correlated increases in the fruit-yield of trees. Thus, experiencing positive prediction errors at one fruit tree ought to inflate our value estimates of nearby fruit trees. Mood provides a possible mechanism whereby this might be achieved: because positive prediction errors create positive mood, which in turn leads to increased value estimates of other trees, we can generalize our learning about one tree to all others. In effect this gives us a way to learn about latent causes, such that changes in specific exemplars are correctly attributed to changes in their underlying causes, permitting updates to other variables dependent upon the same latent cause [195]. Whether mood provides additional inferential power above and beyond good generalization of this kind is unclear.

A more persuasive argument also made by Eldar et al is for generalization in the temporal domain. If changes in value have some underlying trajectory, biasing future prediction of reward in terms of past prediction errors will produce faster learning. By essentially adding a term describing past prediction errors into the current update, we take into account the fact that changes in value over time typically exhibit autocorrelation: things that are getting better tend to keep getting better. Interestingly, autocorrelation of value between adjacent states has proved a headache in neural network models of Q-learning [11], prompting elaborate architecture to overcome the instability that results.

Further consideration of learning in neural networks also suggests a further possibility not discussed by Eldar et al. Modern techniques for training networks using back-propagation [196] increasingly rely upon a measure of momentum, which incorporates gradients on the previous timestep into the current update [197]. Although space precludes a detailed discussion of the benefits of this approach here, suffice to say that is has proved a major accelerant in the training of deep neural networks. The intuition is that noise in individual updates can be overcome by essentially averaging adjacent update terms, to discern the 'true' gradient by which network weights can be optimized (see http://www.willamette.edu/~gorr/classes/cs449/momrate.html). This suggests a possible role for mood in smoothing out rapid fluctuations in prediction error to provide a more robust estimate of environmental trajectory (whether circumstances are getting better or worse).

Current models thus use parameterizations from computational models derived from reinforcement learning to describe how emotion is evoked by and might shape learning processes. In doing so, they have the potential to explain the function of emotion, which remains difficult to pin down. In **Chapter 2** we expand upon this approach, using RL-type models to link emotional and decisional processes, whilst in **Chapter 3** we turn to Bayesian models to understanding emotional and learning dynamics under uncertainty.

1.5.4 Emotional events can affect the balance between learning systems

As reviewed earlier, model-based and model-free control of behaviour appear to be subserved by largely distinct neural circuits [93], which operate simultaneously in the human brain [12,106,109,198]. One prominent line of research is into factors which can shift the balance between these two systems, which seems to be affected by psychiatric conditions such as addiction [110] and Obsessive Compulsive Disorder (OCD) [199]. Accumulating evidence over the past decade suggests that acutely stressful events can temporarily shift the balance between model-based and model-free systems.

A series of studies from Schwabe have provided compelling evidence that exposure to a single stressful event – immersion of the hand in very cold water, referred to as the Cold Pressor Test (CPT) – produces a shift towards habitual learning [200], as assessed by an increased resistance to devaluation procedures. This effect is dependent upon both glucocorticoid and noradrenergic activity and involves reduced recruitment of goal-directed structures such as the medial PFC [201]. A similar recent study using the model-based / model-free framework reached similar conclusions, additionally noting that the impact of stress appeared to be attenuated with increasing working memory capacity [202]. This chimes with evidence in rodents, in which prolonged stress leads to synaptic remodelling consistent with reduced function of circuits associated with control and increased activity in model-free systems [203].

As reviewed earlier, Pavlovian learning promotes approach to rewarding stimuli and avoidance of punishing stimuli regardless of instrumental contingency [102,204], resulting in an interaction

between valence and action in learning [99]. Thus, at least three learning systems* contribute (at times conflicting) input into choice [100]. In **Chapter 3** we ask whether stress also affects Pavlovian contributions to instrumental learning. This is implied by the idea that stress prompts a reorganisation of cognition towards simpler, faster algorithms [208]. Furthermore, dopamine, release of which increases during stress [209], appears to enhance the influence of Pavlovian value estimates upon choice [210,211]. Finally, stress increases the impact of palatability upon food choice [212], in precisely the manner that one would predict from an increased influence of Pavlovian control [100].

1.5.5 Summary and relevance

Emotion can be described as a feedback mechanism that affects ongoing learning and future decision-making. Models of emotion suggest that current emotions reflect whether progress towards a goal is better or worse than expected. Once conceptualization is within the framework of reinforcement learning, with the empirical finding that emotion reflects prediction errors. This has led to the idea that emotions help optimize reinforcement learning in situations where the reward function exhibits spatial or temporal correlations. Emotion might provide inputs both to model-free and model-based systems, manifesting as a bias in value estimation or a distinct state to be avoided or pursued.

In **Chapter 2,** we extend a prediction-error account of emotion into the social domain, and demonstrate a clear relationship between social effects in emotion and subsequent social model-based decision making. In **Chapter 3** we examine the ability of a negative emotion (stress) to affect instrumental and Pavlovian learning processes. This leads to **Chapter 4**, in which we extend prediction-error models of emotion into a Bayesian context, demonstrating that model estimates of uncertainty can explain variation in subjective and physiological stress responses and showing how experience of stress shapes expectations of environmental uncertainty.

---

* The possibility of episodic control, in which memories of distinct episodes are used to guide decisions, has been discussed [205] but received little empirical attention (but see [206] and [207])

1.6 References

1. Wolpert, D. M. The real reason for brains. *Ted Global 2011* (2011). Available at: https://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains/transcript?language=en. (Accessed: 17 October 2016)
2. Thorndike, E. L. The Law of Effect. *The American Journal of Psychology* **39,** 212–222 (1927).
3. Vlaev, I., Chater, N., Stewart, N. & Brown, G. D. A. Does the brain calculate value? *Trends Cogn. Sci. (Regul. Ed.)* **15,** 546–554 (2011).
4. Padoa-Schioppa, C. Commentary: Utility-free heuristic models of two-option choice can mimic predictions of utility-stage models under many conditions. *Frontiers in Neuroscience* **9,** (2015).
5. Kahneman, D. & Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society* 263–291 (1979).
6. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction*. (MIT Press, Cambridge, 1998).
7. Winkielman, P., Knutson, B. & Trujillo, J. Affective Influence on Judgments and Decisions: Moving Towards Core Mechanisms. *Review of General Psychology* **11,** 179–192 (2007).
8. Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience* **33,** 269–298 (2010).
9. Gagliano, M., Renton, M., Depczynski, M. & Mancuso, S. Experience teaches plants to learn faster and forget slower in environments where it matters. *Oecologia* **175,** 63–72 (2014).
10. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* (2012).
11. Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518,** 529–533 (2015).
12. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience* **8,** 1704–1711 (2005).
13. Dennett, D. in *Explaining Consciousness–The "Hard Problem*", ed. Shear, J. (Bradford Books, 2000).
14. Cox, J. C., Sadiraj, V. & Vogt, B. On the empirical relevance of St. Petersburg Lotteries. *Economic Bulletin* (2011).
15. Tversky, A. & Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **185,** 1124–1131 (1974).
16. Simon, H. A. *Administrative behavior*. (New York: The Free Press, 1965).
17. Kahneman, D., Knetsch, J. L. & Thaler, R. H. Fairness and the assumptions of economics. *Journal of business* S285–S300 (1986).
18. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 817–868 (1999).
19. Nishi, A., Shirado, H., Rand, D. G. & Christakis, N. A. Inequality and visibility of wealth in experimental social networks. *Nature* **526,** 426–429 (2015).
20. Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. Egalitarian motives in humans. *Nature* **446,** 794–796 (2007).
21. Dawes, C. T. *et al.* Neural basis of egalitarian behavior. *Proceedings of the National Academy of Sciences* **109,** 6479–6483 (2012).
22. Chang, S. W. C. *et al.* Neuroethology of primate social behavior. *Proceedings of the National Academy of Sciences* **110,** 10387–10394 (2013).
23. Brosnan, S. F. & de Waal, F. B. M. Evolution of responses to (un)fairness. *Science* **346,** 1251776 (2014).
24. Fehr, E. & Camerer, C. F. Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci. (Regul. Ed.)* **11,** 419–427 (2007).
25. O'Doherty, J. P. The problem with value. *Neurosci Biobehav Rev* **43,** 259–268 (2014).
26. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76,** 412–427 (2013).

27. Tom, S. M., Fox, C. R., Trepel, C. & Poldrack, R. A. The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science* **315,** 515–518 (2007).
28. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience* **10,** 1625–1633 (2007).
29. De Martino, B., Kumaran, D., Ben Seymour & Dolan, R. J. Frames, biases, and rational decision-making in the human brain. *Science* **313,** 684–687 (2006).
30. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27,** 9984–9988 (2007).
31. FitzGerald, T. H. B., Seymour, B. & Dolan, R. J. The role of human orbitofrontal cortex in value comparison for incommensurable objects. *J. Neurosci.* **29,** 8388–8395 (2009).
32. Howard, J. D., Gottfried, J. A., Tobler, P. N. & Kahnt, T. Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proceedings of the National Academy of Sciences* **112,** 5195–5200 (2015).
33. Azim, E., Mobbs, D., Jo, B., Menon, V. & Reiss, A. L. Sex differences in brain activation elicited by humor. *Proceedings of the National Academy of Sciences* **102,** 16496–16501 (2005).
34. Aharon, I. *et al.* Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* **32,** 537–551 (2001).
35. Tricomi, E., Rangel, A., Camerer, C. F. & O'Doherty, J. P. Neural evidence for inequality-averse social preferences. *Nature* **463,** 1089–1091 (2010).
36. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441,** 223–226 (2006).
37. Padoa-Schioppa, C. Neurobiology of economic choice: a good-based model. *Annual Review of Neuroscience* **34,** 333–359 (2011).
38. Kennerley, S. W., Behrens, T. E. J. & Wallis, J. D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience* **14,** 1581–1589 (2011).
39. Hosokawa, T., Kennerley, S. W., Sloan, J. & Wallis, J. D. Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *J. Neurosci.* **33,** 17385–17397 (2013).
40. Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience* **9,** 545–556 (2008).
41. Montague, P. R. & Berns, G. S. Neural Economics and the Biological Substrates of Valuation. *Neuron* **36,** 265–284 (2002).
42. Fellows, L. K. Deciding how to decide: ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. *Brain* **129,** 944–952 (2006).
43. Wang, X.-J. Decision making in recurrent neuronal circuits. *Neuron* **60,** 215–234 (2008).
44. Hunt, L. *et al.* Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience* **15,** 470–6– S1–3 (2012).
45. Marblestone, A. H. & Wayne, G. Toward an integration of deep learning and neuroscience. *Front. Comput. Neurosci.* (2016).
46. deCharms, R. C. & Zador, A. Neural representation and the cortical code. *Annu. Rev. Neurosci.* **23,** 613–647 (2000).
47. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature Neuroscience* **10,** 1214–1221 (2007).
48. Glautier, S. Revisiting the learning curve (once again). *Front Psychol* **4,** (2013).
49. Kamin, L. J. Predictability, surprise, attention, and conditioning. *Punishment and aversive behaviour* 279–296 (1969).
50. Rescorla, R. A. Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning and Motivation* (1970).
51. Takahashi, Y. K. *et al.* Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron* **80,** 507–518 (2013).
52. Sutton, R. S. & Barto, A. G. Toward a modern theory of adaptive networks: expectation and prediction. *Psychological Review* **88,** 135–170 (1981).

53. Schultz, W., Dayan, P. & al, E. A neural substrate of prediction and reward. *Science* (1997).
54. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-Brain Mapping of Direct Inputs to Midbrain Dopamine Neurons. *Neuron* **74,** 858–873 (2012).
55. Haber, S. N. & Fudge, J. L. The Primate Substantia Nigra and VTA: Integrative Circuitry and Function. *Crit Rev Neurobiol* **11,** 323–342 (1997).
56. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience* **16,** 966–973 (2013).
57. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482,** 85–88 (2012).
58. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47,** 129–141 (2005).
59. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412,** 43–48 (2001).
60. Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299,** 1898–1902 (2003).
61. Fiorillo, C. D. Two Dimensions of Value: Dopamine Neurons Represent Reward But Not Aversiveness. *Science* **341,** 546–549 (2013).
62. Tobler, P. N., Dickinson, A. & Schultz, W. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J. Neurosci.* **23,** 10402–10410 (2003).
63. Seymour, B. *et al.* Temporal difference models describe higher-order learning in humans. *Nature* **429,** 664–667 (2004).
64. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442,** 1042–1045 (2006).
65. D'Ardenne, K., McClure, S. M., Nystrom, L. E. & Cohen, J. D. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* **319,** 1264–1267 (2008).
66. O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38,** 329–337 (2003).
67. Rutledge, R. B., Dean, M., Caplin, A. & Glimcher, P. W. Testing the reward prediction error hypothesis with an axiomatic model. *J. Neurosci.* **30,** 13525–13536 (2010).
68. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nature Neuroscience* **18,** 1213–1225 (2015).
69. Chang, C. Y. *et al.* Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nature Neuroscience* **19,** 111–116 (2015).
70. Eshel, N. *et al.* Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525,** 243–246 (2015).
71. Matsumoto, M. & Hikosaka, O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459,** 837–841 (2009).
72. Lammel, S. *et al.* Input-specific control of reward and aversion in the ventral tegmental area. *Nature* **491,** 1–8 (2012).
73. Lammel, S., Ion, D., Roeper, J. & Malenka, R. Projection-Specific Modulation of Dopamine Neuron Synapses by Aversive and Rewarding Stimuli. *Neuron* **70,** 855–862 (2011).
74. Matsumoto, H., Tian, J., Uchida, N. & Watabe-Uchida, M. Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. *eLife Sciences* **5,** (2016).
75. Flagel, S. B. *et al.* A selective role for dopamine in stimulus-reward learning. *Nature* **469,** 53–57 (2011).
76. Cannon, C. M. & Palmiter, R. D. Reward without dopamine. *J. Neurosci.* **23,** 10827–10831 (2003).
77. Adamantidis, A. R. *et al.* Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *J. Neurosci.* **31,** 10829–10835 (2011).
78. Bromberg-Martin, E. S., Matsumoto, M. & Hikosaka, O. Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* **68,** 815–834 (2010).
79. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. *Nature*

Reviews Neuroscience **17,** 183–195 (2016).
80. Thorndike, E. L. A proof of the Law of Effect. *Science* **77,** 173–175 (1933).
81. Peterson, G. B. A day of great illumination: BF Skinner's discovery of shaping. *J Exp Anal Behav* (2004).
82. Watkins, C. & Dayan, P. Q-learning. *Machine Learning* **8,** 279–292 (1992).
83. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* **91,** 1402–1412 (2016).
84. Doya, K. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology* **10,** 732–739 (2000).
85. Mishkin, M., Malamut, B. & Bachevalier, J. in Neurobiology of human learning and memory (1984).
86. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310,** 1337–1340 (2005).
87. Lau, B. & Glimcher, P. W. Value Representations in the Primate Striatum during Matching Behavior. *Neuron* **58,** 451–463 (2008).
88. Seo, M., Lee, E. & Averbeck, B. B. Action selection and action value in frontal-striatal circuits. *Neuron* **74,** 947–960 (2012).
89. Fitzgerald, T. H. B., Friston, K. J. & dolan, R. J. Action-Specific Value Signals in Reward-Related Regions of the Human Brain. *J. Neurosci.* **32,** 16417–16423 (2012).
90. Haber, S. N. The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* **26,** 317–330 (2003).
91. O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304,** 452–454 (2004).
92. Frank, M. J. Computational models of motivated action selection in corticostriatal circuits. *Current Opinion in Neurobiology* **21,** 381–386 (2011).
93. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80,** 312–325 (2013).
94. Niv, Y. Reinforcement learning in the brain. *Journal of Mathematical Psychology* (2009).
95. Rutledge, R. B. *et al.* Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *J. Neurosci.* **29,** 15104–15114 (2009).
96. de Berker, A. O. & Rutledge, R. B. A role for the human substantia nigra in reinforcement learning. *J. Neurosci.* **34,** 12947–12949 (2014).
97. Ramayya, A. G., Misra, A., Baltuch, G. H. & Kahana, M. J. Microstimulation of the human substantia nigra alters reinforcement learning. *J. Neurosci.* **34,** 6887–6895 (2014).
98. Huys, Q. J. M. & Dayan, P. A Bayesian formulation of behavioral control. *Cognition* **113,** 314–328 (2009).
99. Guitart-Masip, M., Düzel, E., Dolan, R. & Dayan, P. Action versus valence in decision making. *Trends Cogn. Sci. (Regul. Ed.)* **18,** 194–202 (2014).
100. Dayan, P., Niv, Y., Seymour, B. & Daw, N. D. The misbehavior of value and the discipline of the will. *Neural networks : the official journal of the International Neural Network Society* **19,** 1153–1160 (2006).
101. Guitart-Masip, M. *et al.* Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage* **62,** 154–166 (2012).
102. Huys, Q. J. M. *et al.* Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLoS Comp Biol* **7,** e1002028 (2011).
103. Tolman, E. C. Cognitive maps in rats and men. *Psychological Review* **55,** 189–208 (1948).
104. Dayan, P. Improving generalization for temporal difference learning: The successor representation. *Neural Comput* (1993).
105. Wunderlich, K., Dayan, P. & Dolan, R. J. Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience* **15,** 786–791 (2012).
106. Daw, N. D., Gershman, S. J., Seymour, B. & Dayan, P. Model-based influences on humans' choices and striatal prediction errors. *Neuron* (2011).
107. Wunderlich, K., Smittenaar, P. & Dolan, R. J. Dopamine enhances model-based over model-free

choice behavior. *Neuron* **75,** 418–424 (2012).
108. Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D. & Dolan, R. J. Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* **80,** 914–919 (2013).
109. Lee, S. W., Shimojo, S. & O'Doherty, J. P. Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron* **81,** 687–699 (2014).
110. Everitt, B. J. & Robbins, T. W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience* **8,** 1481–1489 (2005).
111. Adams, C. D. & Dickinson, A. Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology. B* **33,** 109–121 (1981).
112. Dickinson, A., Balleine, B., Watt, A., Gonzalez, F. & Boakes, R. A. Motivational control after extended instrumental training. *Animal Learning & Behavior* **23,** 197–206 (1995).
113. Tricomi, E., Balleine, B. W. & O'Doherty, J. P. A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* **29,** 2225–2232 (2009).
114. Zapata, A., Minney, V. L. & Shippenberg, T. S. Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. *J. Neurosci.* **30,** 15457–15463 (2010).
115. Nelson, A. & Killcross, S. Amphetamine exposure enhances habit formation. *J. Neurosci.* **26,** 3805–3812 (2006).
116. Izquierdo, A., Suda, R. K. & Murray, E. A. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* **24,** 7540–7548 (2004).
117. Jones, J. L. *et al.* Orbitofrontal Cortex Supports Behavior and Learning Using Inferred But Not Cached Values. *Science* **338,** 953–956 (2012).
118. Yin, H. H., Ostlund, S. B., Knowlton, B. J. & Balleine, B. W. The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.* **22,** 513–523 (2005).
119. Friedel, E. *et al.* Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Front Hum Neurosci* **8,** 587 (2014).
120. MacKay, D. *Information theory, inference and learning algorithms*. (Cambridge University Press, 2003).
121. O'Reilly, J. X., Jbabdi, S. & Behrens, T. E. J. How can a Bayesian approach inform neuroscience? *European Journal of Neuroscience* **35,** 1169–1179 (2012).
122. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences* **111,** 12252–12257 (2014).
123. Hart, A. S., Rutledge, R. B., Glimcher, P. W. & Phillips, P. E. M. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J. Neurosci.* **34,** 698–704 (2014).
124. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46,** 681–692 (2005).
125. Preuschoff, K. & Bossaerts, P. Adding Prediction Risk to the Theory of Rep=0.98ward Learning. *Annals of the New York Academy of Sciences* **1104,** 135–146 (2007).
126. Daw, N. D. in (Neuroeconomics: Decision-Making and the Brain, 2013).
127. Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. Probabilistic brains: knowns and unknowns. *Nature Neuroscience* **16,** 1170–1178 (2013).
128. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415,** 429–433 (2002).
129. Battaglia, P. W., Jacobs, R. A. & Aslin, R. N. Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A Opt Image Sci Vis* **20,** 1391–1397 (2003).
130. Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* (2004).
131. Bahrami, B. *et al.* Optimally Interacting Minds. *Science* **329,** 1081–1085 (2010).
132. Kording, K. Decision Theory: What 'Should' the Nervous System Do? *Science* **318,** 606–610 (2007).
133. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A Bayesian foundation for individual

learning under uncertainty. *Front Hum Neurosci* **5,** 39 (2011).
134. Payzan-LeNestour, E. & Bossaerts, P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comp Biol* **7,** e1001048 (2011).
135. Iglesias, S. *et al.* Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* **80,** 519–530 (2013).
136. Diaconescu, A. O. *et al.* Inferring on the Intentions of Others by Hierarchical Bayesian Learning. *PLoS Comp Biol* **10,** (2014).
137. Fiser, J., Berkes, P., Orbán, G. & Lengyel, M. Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci. (Regul. Ed.)* **14,** 119–130 (2010).
138. Bach, D. R. & Dolan, R. J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience* **13,** 572–586 (2012).
139. O'Neill, M. & Schultz, W. Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron* **68,** 789–800 (2010).
140. Preuschoff, K., Bossaerts, P. & Quartz, S. R. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* **51,** 381–390 (2006).
141. Wright, N. D., Symmonds, M. & Dolan, R. J. Distinct encoding of risk and value in economic choice between multiple risky options. *Neuroimage* **81,** 431–440 (2013).
142. McCoy, A. N. & Platt, M. L. Risk-sensitive neurons in macaque posterior cingulate cortex. *Nature Neuroscience* **8,** 1220–1227 (2005).
143. Hsu, M., Bhatt, M., Adolphs, R., Tranel, D. & Camerer, C. F. Neural systems responding to degrees of uncertainty in human decision-making. *Science* **310,** 1680–1683 (2005).
144. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455,** 227–231 (2008).
145. Critchley, H. D., Mathias, C. J. & Dolan, R. J. Neural Activity in the Human Brain Relating to Uncertainty and Arousal during Anticipation. *Neuron* **29,** 537–545 (2001).
146. Monosov, I. E. & Hikosaka, O. Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nature Neuroscience* **16,** 756–762 (2013).
147. Ledbetter, N. M., Chen, C. D. & Monosov, I. E. Multiple Mechanisms for Processing Reward Uncertainty in the Primate Basal Forebrain. *J. Neurosci.* **36,** 7852–7864 (2016).
148. Monosov, I. E., Leopold, D. A. & Hikosaka, O. Neurons in the Primate Medial Basal Forebrain Signal Combined Information about Reward Uncertainty, Value, and Punishment Anticipation. *J. Neurosci.* **35,** 7443–7459 (2015).
149. Tobler, P. N., O'Doherty, J. P., Dolan, R. J. & Schultz, W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* **97,** 1621–1632 (2007).
150. Platt, M. L. & Huettel, S. A. Risky business: the neuroeconomics of decision making under uncertainty. *Nature Neuroscience* **11,** 398–403 (2008).
151. Ogawa, M. *et al.* Risk-responsive orbitofrontal neurons track acquired salience. *Neuron* **77,** 251–258 (2013).
152. Pearce, J. M. & Hall, G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review* **87,** 532–552 (1980).
153. Esber, G. R. & Haselgrove, M. Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. *Proceedings of the Royal Society B: Biological Sciences* **278,** 2553–2561 (2011).
154. Doya, K. Modulators of decision making. *Nature Neuroscience* **11,** 410–416 (2008).
155. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience* **15,** 1040–1046 (2012).
156. Preuschoff, K., 't Hart, B. M. & Einhäuser, W. Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience* **5,** 115 (2011).
157. Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X. & Bishop, S. J. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience* **18,** 590–596

(2015).
158. Dayan, P. Twenty-five lessons from computational neuromodulation. *Neuron* **76,** 240–256 (2012).
159. Jimenez Rezende, D. & Gerstner, W. Stochastic variational learning in recurrent spiking networks. *Front. Comput. Neurosci.* **8,** (2014).
160. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. *arXiv.org* (2013).
161. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *arXiv.org* (2014).
162. Zemel, R. S., Dayan, P. & Pouget, A. Probabilistic interpretation of population codes. *Neural Comput* **10,** 403–430 (1998).
163. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nature Neuroscience* **9,** 1432–1438 (2006).
164. Beck, J. M. *et al.* Probabilistic population codes for Bayesian decision making. *Neuron* **60,** 1142–1152 (2008).
165. van Bergen, R. S., Ma, W. J., Pratte, M. S. & Jehee, J. F. M. Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience* **18,** 1728–1730 (2015).
166. Berkes, P., Orban, G., Lengyel, M. & fiser, J. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331,** 83 (2011).
167. Okun, M. *et al.* Population Rate Dynamics and Multineuron Firing Patterns in Sensory Cortex. *J. Neurosci.* **32,** 17108–17119 (2012).
168. Orbán, G., Berkes, P., Fiser, J. & Lengyel, M. Neural Variability and Sampling-Based Probabilistic Representations in the Visual Cortex. *Neuron* **92,** 530–543 (2016).
169. Murray, J. D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience* **17,** 1661–1663 (2014).
170. Cavanagh, S. E., Wallis, J. D., Kennerley, S. W. & Hunt, L. T. Autocorrelation structure at rest predicts value correlates of single neurons during reward-guided choice. *eLife Sciences* **5,** 1–39 (2016).
171. Sequeira, P., Melo, F. S. & Paiva, A. Emotion-based intrinsic motivation for reinforcement learning agents. *International Conference on Affective Computing and Intelligent Interaction* 326-336 (2011).
172. Darwin, C. *The expression of the emotions in man and animals*. (Oxford University Press, 1998).
173. James, W. II.—What is an emotion? *Mind* (1884).
174. Friedman, B. H. Feelings and the body: the Jamesian perspective on autonomic specificity of emotion. *Biol Psychol* **84,** 383–393 (2010).
175. Schachter, S. in *Advances in Experimental Social Psychology* **1,** 49–80 (Elsevier, 1964).
176. Oatley, K. & Johnson-Laird, P. N. Cognitive approaches to emotions. *Trends Cogn. Sci. (Regul. Ed.)* **18,** 134–140 (2014).
177. Ekman, P., Levenson, R. & Friesen, W. Autonomic nervous system activity distinguishes among emotions. *Science* **221,** 1208–1210 (1983).
178. Russell, J. A. Emotion, core affect, and psychological construction. *Cognition & Emotion* **23,** 1259–1283 (2009).
179. Baumeister, R. F., Vohs, K. D., DeWall, C. N. & Zhang, L. How emotion shapes behavior: feedback, anticipation, and reflection, rather than direct causation. *Pers Soc Psychol Rev* **11,** 167–203 (2007).
180. Cahill, L., Prins, B., Weber, M. & McGaugh, J. L. Beta-adrenergic activation and memory for emotional events. *Nature* **371,** 702–704 (1994).
181. Strange, B. A. & Dolan, R. β-Adrenergic modulation of emotional memory-evoked human amygdala and hippocampal responses. *Proceedings of the National Academy of Sciences* **101,** 11454–11458 (2004).
182. Manucia, G. K., Baumann, D. J. & Cialdini, R. B. Mood influences on helping: Direct effects or side effects? *Journal of Personality and Social Psychology* **46,** 357–364 (1984).
183. Benjamin, D. J., Heffetz, O., Kimball, M. S. & Rees-Jones, A. What Do You Think Would Make You Happier? What Do You Think You Would Choose? *American Economic Review* **102,** 2083–2110 (2012).

184. Gilbert, D. T. & Wilson, T. D. Why the brain talks to itself: sources of error in emotional prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364,** 1335–1341 (2009).
185. Blanco, M., Engelmann, D. & Normann, H. T. A within-subject analysis of other-regarding preferences. *Games and Economic Behavior* **72,** 321–338 (2011).
186. Carver, C. S. & Scheier, M. F. Origins and functions of positive and negative affect: A control-process view. *Psychological Review* **97,** 19–35 (1990).
187. Carver, C. Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect. *Cognition & Emotion* **17,** 241–261 (2003).
188. Oatley, K. & Johnson-Laird, P. N. Towards a Cognitive Theory of Emotions. *Cognition & Emotion* **1,** 29–50 (1987).
189. Toda, M. The design of a fungus-eater: A model of human behavior in an unsophisticated environment. *Behav Sci* **7,** 164–183 (1962).
190. Singer, T., Critchley, H. D. & Preuschoff, K. A common role of insula in feelings, empathy and uncertainty. *Trends Cogn. Sci. (Regul. Ed.)* **13,** 334–340 (2009).
191. Gu, X., Hof, P. R., Friston, K. J. & Fan, J. Anterior insular cortex and emotional awareness. *The Journal of comparative neurology* **521,** 3371–3388 (2013).
192. Eldar, E. & Niv, Y. Interaction between emotional state and learning underlies mood instability. *Nat Comms* **6,** 6149 (2015).
193. Mayer, J. D., Gaschke, Y. N., Braverman, D. L. & Evans, T. W. Mood-congruent judgment is a general effect. *Journal of Personality and Social Psychology* **63,** 119–132 (1992).
194. Aïte, A. *et al.* Impact of emotional context congruency on decision making under ambiguity. *Emotion* **13,** 177–182 (2013).
195. Gershman, S. J. & Niv, Y. Learning latent structure: carving nature at its joints. *Current Opinion in Neurobiology* **20,** 251–256 (2010).
196. Rumelhart, D. E., Hinton, G. E. & Williams, D. Learning representations by back-propagating errors. *Nature* **323,** 333–536 (1986).
197. Sutskever, I., Martens, J., Dahl, G. E. & Hinton, G. E. On the importance of initialization and momentum in deep learning. *ICML (3)* (2013).
198. Glaascher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66,** 585–595 (2010).
199. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife Sciences* **5,** (2016).
200. Schwabe, L. & Wolf, O. T. Stress prompts habit behavior in humans. *J. Neurosci.* **29,** 7191–7198 (2009).
201. Schwabe, L., Tegenthoff, M., Hoffken, O. & Wolf, O. T. Concurrent glucocorticoid and noradrenergic activity shifts instrumental behavior from goal-directed to habitual control. *J. Neurosci.* **30,** 8190–8196 (2010).
202. Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences* **110,** 20941–20946 (2013).
203. Dias-Ferreira, E. *et al.* Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* **325,** 621–625 (2009).
204. Hershberger, W. A. An approach through the looking-glass. *Animal Learning & Behavior* **14,** 443–451 (1986).
205. Lengyel, M. & Dayan, P. Hippocampal Contributions to Control: The Third Way. *Advances in Neural Information Processing Systems* (2007).
206. Wimmer, G. E., Braun, E. K., Daw, N. D. & Shohamy, D. Episodic memory encoding interferes with reward learning and decreases striatal prediction errors. *J. Neurosci.* **34,** 14901–14912 (2014).
207. Blundell, C., Uria, B., Pritzel, A., Li, Y. & Ruderman, A. Model-Free Episodic Control. *arXiv.org* (2016).
208. Steinhauser, M., Maier, M. & Hübner, R. Cognitive control under stress: how stress affects

strategies of task-set reconfiguration. *Psychol Sci* **18,** 540–545 (2007).
209. Pruessner, J. C. Dopamine Release in Response to a Psychological Stress in Humans and Its Relationship to Early Life Maternal Care: A Positron Emission Tomography Study Using [11C]Raclopride. *J. Neurosci.* **24,** 2825–2831 (2004).
210. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. Dopaminergic Modulation of Decision Making and Subjective Well-Being. *J. Neurosci.* **35,** 9811–9822 (2015).
211. Rigoli, F. *et al.* Dopamine Increases a Value-Independent Gambling Propensity. *Neuropsychopharmacology* **41,** 2658–2667 (2016).
212. Maier, S. U., Makwana, A. B. & Hare, T. A. Acute Stress Impairs Self-Control in Goal-Directed Choice by Altering Multiple Functional Connections within the Brain's Decision Circuits. *Neuron* **87,** 621–631 (2015).

Chapter 2: The social contingency of momentary subjective well-being

Previously published as:

Rutledge R. B.*, de Berker A. O.*, Espenhahn S., Dayan P. & Dolan R. J. The social contingency of momentary subjective well-being. *Nature Communications* **7**, 11825 (2016).

* Denotes shared first-authorship

2.1 Abstract

Here we use the reinforcement learning models introduced in the previous chapter, demonstrating how they can contribute to our understanding of emotional state. We first relate fluctuations in emotional state in participants playing a gambling task to the expectations and prediction errors they experience, replicating previous work. We then develop the model further to account for the experiences of other players in the environment. This gives us the opportunity to examine how our emotional characterization of social preferences relates to choice in a separate social decision task. We find that a participant's subjective emotional state reflects impact of rewards and prediction errors they themselves receive, and the rewards received by a social partner. Unequal outcomes, whether advantageous or disadvantageous, reduce average momentary happiness. Furthermore, we show that the relative impacts of advantageous and disadvantageous inequality on momentary happiness at the individual level predict a subject's generosity in a separate dictator game. These findings demonstrate a powerful social influence upon subjective emotional state, with emotional reactivity to inequality predictive of altruism in an independent task domain. This provides a quantitative account of the often elusive link between emotional and decisional processes.

2.2 Introduction

Subjective well-being is a key index of quality of life[1,2], prompting policies aimed at increasing it[3]. However, maximizing wealth is not an effective way of maximizing well-being, as the coupling between the two is often relatively weak[4-7] (although see ref. 8). Social comparison has been suggested as an important mediator of the relationship between wealth and well-being, with relative as opposed to absolute wealth exerting a substantial influence on well-being[9-11].

Social comparison is also increasingly acknowledged as being relevant to economic behaviour, as described in section 1.2.1. Aversion to advantageous and disadvantageous inequality is suggested to underlie altruistic behaviour[12,13]. However, it is unknown whether variance in the emotional impact of inequality on well-being relates to the hitherto poorly understood heterogeneity observed in altruistic behavior[14]. Here we address these issues at the level of the individual by examining the impact of social comparison on momentary subjective well-being. We show that, on average, the impact of inequality is to attenuate momentary happiness. Furthermore, the relative emotional impact of advantageous and disadvantageous inequality predicts altruistic behavior at the level of individuals.

We previously quantified the relationship between rewards and momentary happiness using a probabilistic reward task, showing that momentary subjective well-being depends on the cumulative impact of recent expectations and the reinforcement prediction errors (RPEs) that arise from these expectations[15] (see section 1.5.1). RPEs, the difference between experienced and expected outcomes, are thought to be encoded in the firing pattern of dopamine neurons[16,17] (see section 1.3.1.3). In keeping with this, we observed that changes in subjective well-being were coupled to reward-related neural responses in the striatum, an area with rich dopaminergic innervation[15]. A dopaminergic mediation of this effect was also suggested by the observation that pharmacologically boosting dopamine increased well-being related to reward receipt[18].

Here, we exploit our previously established computational approach to study how inequality, and putative emotional responses to inequality, impact on one important component of well-being. Our approach was motivated by prior observations that striatal neural responses can also reflect the rewards received by others[19-21]. This led us to predict that rewards received by

another person would impact participants' momentary subjective well-being according to their individual social preferences. We also predicted that individual emotional reactivity to social outcomes might relate to heterogeneity in altruistic choice, something that has been difficult to explain using standard economic approaches[14].

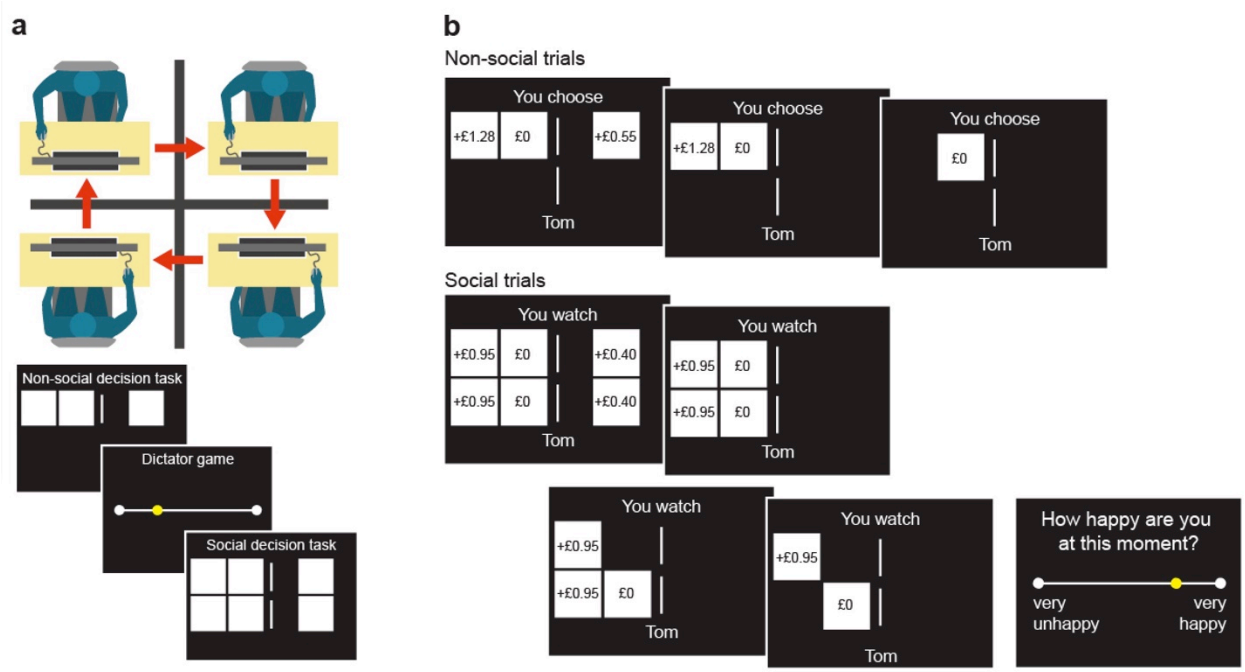


Figure 2.1 | Experimental design (A) Four participants were introduced to each other and seated in separate rooms to perform three tasks. In the non-social decision task, subjects (*n*=47) chose between safe options and risky gambles with equal probabilities of two outcomes. In the dictator game, subjects decided how to split an endowment between themselves and another player. **(B)** The social decision task consisted of non-social and social trials. In non-social trials, choice outcomes (here £0) did not affect partner earnings. In social trials, subjects were told that they were observing choices made by their partner in the non-social decision task. When their partner chose to gamble, the subject received an equivalent but independent gamble. The subject's outcome was revealed first (here gaining £0.95), followed by the partner's outcome (here £0). After every 2-3 trials, subjects were asked to report their current level of happiness.

Our results show that a subject's subjective emotional state reflects rewards received by a social partner. Advantageous and disadvantageous inequality both reduce momentary happiness on average. Furthermore, we use computational modeling to show that the relative emotional impacts of advantageous and disadvantageous inequality predict a subject's generosity in a

separate dictator game, suggesting that variability in the emotional impact of inequality on well-being can explain heterogeneity in altruistic behaviour.

2.3 Methods

2.3.1 Participants

Forty-seven healthy subjects took part in the experiment (age range 18-39, 22 male), using two slightly different procedures (n=22 and n=25). Same-gendered subjects who did not know each other were tested in groups of four and confederates were used when one of the scheduled subjects was absent (Figure 2.1A). The experimenter asked subjects to introduce themselves to the other members of the group before seating them in four separate rooms. All subjects gave informed consent and the Research Ethics Committee of University College London approved all studies.

2.3.2 Experimental procedure

2.3.2.1 Solo decision task

Stimuli were presented using Cogent 2000 (Wellcome Trust Centre for Neuroimaging) in MATLAB (MathWorks, Inc.). First, subjects completed a non-social decision task with 140 trials (Figure 2.1B). Subjects completed instructions and a practice session before the task. On each trial, subjects made a choice between a safe and a risky option which was resolved after a 2.5 s delay period[15,18,22]. Subjects faced Gain trials (certain gain vs a gamble to gain a larger amount or zero), Mixed trials (zero vs a gamble to gain an amount or lose an amount), and Loss trials (certain loss vs a gamble to lose a larger amount or zero). Options presented in the task were similar to those used in a previous study[15]. Subjects had 5 s to make their decision and otherwise received the worst outcome from the gamble. The position of safe and risky options was left-right reversed every 10 trials.

To familiarize subjects with answering questions about their subjective emotional state, during the non-social decision task we used the same key measure obtained in the social decision task, asking subjects the question “How happy are you at this moment?” at the start of the task and after every 10 trials. The left side of the line was marked “very unhappy” and the right side of the line was marked “very happy”. Subjects moved a cursor to indicate their current subjective state. The cursor always started at the midpoint. Subjects had a 5 s time limit to make their responses

and the current cursor position was entered as the response if they did not respond within the time limit. The average decision time was 1.4 s and the average rating time was 2.0 s in the non-social decision task. Total task earnings were not revealed to subjects during the experiment; subjects were told that the computer would track all of their earnings and they would be told the combined total for all tasks and receive those earnings at the end of the experiment. Although this task only included 15 happiness ratings, we fitted our pre-existing non-social happiness model to z-scored happiness ratings and found, as expected, that **CR**, **EV**, and **RPE** weights were on average positive (means: **CR**=0.83, **EV**=0.62, **RPE**=1.15; all $Z>4$, $P<0.001$). The forgetting factor was 0.65±0.31 (mean±SD).

2.3.2.2 Dictator game

Second, subjects completed a dictator game in which they were endowed with an amount of money and tasked with splitting the money between themselves and a named partner (Figure 2.1B). They were told that the split was anonymous and would be added to the partner's total earnings without the partner's knowledge. Subjects had no time limit to make their decisions in the dictator game. The social decision task that immediately followed was with the same named partner. In procedure 1 (n=22), subjects were paired with only one partner, completing a single dictator game with an endowment of £3. In procedure 2 (n=25), subjects were paired sequentially with two partners, completing a dictator game with an endowment of £2 before a social decision task with each partner. We employed this procedure to determine if there was any variability in generosity within subjects that could be exploited to examine differences in emotional reactivity to outcomes received by the two partners. No happiness ratings were collected during this task.

Generosity in the dictator game was almost identical on average between procedure 1 and procedure 2 (20% vs 19%; $Z=0.09$, $P=0.93$). However, generosity in the dictator game in procedure 2 was highly correlated between the two partners (Spearman's $\rho=0.88$, $P<0.001$), suggesting that generosity in this task is stable, at least with unfamiliar partners. To further ascertain if subjects had any preference for one of the partners that might impact generosity, we asked subjects at the end of the experiment which unfamiliar partner they would prefer to have a conversation with. There was no difference in generosity toward preferred and non-preferred partners ($Z= -1.09$, $P=0.27$). Due to the high degree of similarity in generosity across repeated

dictator games in procedure 2, we combined data from the two partners and took the mean of generosity in the two dictator games.

2.3.2.3 Social decision task

Third, subjects completed a social decision task in which on each trial they were again presented with a safe and risky option (Figure 2.1B). The task consisted of non-social and social trials, with the order pseudo-randomized to ensure that there were never more than 2 non-social trials or 4 social trials in a row. In non-social trials, subjects chose as in the non-social decision task and the outcome of decisions was added to their earnings. Subjects had 5 s to make their decision and otherwise received the worst outcome from the gamble. In social trials, they were told that they were observing the choice made by the social partner when that partner earlier completed the non-social decision task. If the partner chose the safe option, that outcome was added to their earnings. If the partner chose the gamble, independent gambles were resolved for the subject and the partner. The subject's outcome was resolved first after a 2.5 s delay period. The partner's outcome was resolved after an additional 2.5 s delay period. Subjects were asked after every 2-3 trials 'How happy are you at this moment?' providing the opportunity to examine the effect of inequality on subjective emotional state.

Decisions in social trials were not in fact made by the partner but were made by the computer in a manner approximating an agent with typical economic preferences. This agent made choices based on the parametric prospect theory model with typical loss aversion (λ=1.35) and typical risk aversion in gains and risk seeking in losses (ρ=0.9). These parameters are similar to average parameter values obtained in a non-social experiment with a similar design[22].

In procedure 1, the social decision task involved 210 trials (70 non-social trials and 140 social trials) including 85 happiness ratings. In procedure 2, each social decision task involved 150 trials (50 non-social trials and 100 social trials) including 61 happiness ratings. The percentage of trials in which subjects chose to gamble was similar in non-social and social decision tasks (median of 54% in both tasks) and similar to the percentage of trials in which the computer partner gambled (median of 55%). In the social decision task, the average decision time in the non-social trials was 1.6 s and the average rating time was 1.7 s. On average, subjects responded within the time-limit in 99% of non-social trials and 96% of rating trials.

2.3.3 Descriptive and model-based analyses

Happiness ratings were z-scored so that subjects with greater variability in their ratings did not disproportionately contribute to results. Due to the non-normality of decisions in the dictator game, with 26 subjects giving either half or nothing (Figure 2.2B), we used non-parametric statistical tests including two-tailed Wilcoxon signed-rank tests, Wilcoxon rank-sum tests, and Spearman's rank correlations. Unless otherwise stated, statistical tests included all 47 subjects.

We modelled momentary happiness using models that assume an exponential decay in the influence of previous events[15,18]. Models were fit to ratings in individual subjects using nonlinear least squares implemented in the optimization toolbox in MATLAB (Mathworks, Inc.). Z-scoring produces ratings with a mean value of 0, eliminating the need for a constant term in the model. The non-social model accounted for at least 10% of the variance in ratings for 45 of 47 subjects. The simple-inequality model included a parameter for the magnitude of the difference in rewards between the two players. The guilt-envy model included parameters for the magnitude of advantageous inequality (guilt) and the magnitude of disadvantageous inequality (envy). These parameters will be negative if either type of inequality reduces well-being. We used Bayesian model comparison to compare models[30,31]. We computed Bayesian Information Criterion (BIC) measures for each individual model fit and summed across subjects. BIC is a measure that quantifies the deviation of the model's predictions from the data. A lower BIC value is therefore preferable. However, BIC also penalizes for the number of parameters, allowing the direct comparison of models with different numbers of parameters. Because the relative BIC value is important, and not the absolute BIC value, we also computed the BIC values relative to the winning model (Table 2.1).

2.4 Results

Our experimental design involved testing subjects in groups of four. Subjects (n=47) were first introduced to each other, then seated in separate rooms and asked to complete three different tasks (Figure 2.1A). The first task was a non-social decision task[15,22,23] in which subjects chose between safe and risky options. Subjects faced Gain trials (certain gain vs a gamble to gain a larger amount or zero), Mixed trials (zero vs a gamble to gain an amount or lose an amount), and Loss trials (certain loss vs a gamble to lose a larger amount or zero). Chosen gambles were resolved after a brief delay and the outcomes of all trials counted toward earnings. The second

task was a standard economic task, the dictator game, in which a subject decided how to split an endowment (either £2 or £3, see Methods) with one of the other players[24]. Importantly, these allocations were private and subjects were told that the monetary split would never be revealed to the other player. At the end of the experiment, subjects learned their total earnings for completing all tasks and were not told if, or how much, any other player had contributed to that total. This design feature, whereby allocations are private, is important because generosity might otherwise reflect a concern for what other players will think of them[25]. Generosity was estimated based on behavior in the dictator game as the percentage of the endowment that subjects allocated to their social partner. This quantity varied between 0% and 50%, consistent with previous research[26]. The third and final task involved social and non-social decision trials, where subjects were again presented with safe and risky options (Figure 2.1B). In the non-social trials, subjects made choices as in the first task. In the social trials, subjects were shown two sets of identical safe and risky options and informed that one set was allocated to them, and the other set corresponded to a trial the partner had previously experienced in the non-social decision task. Subjects were informed that on these trials they could not make a decision for themselves, but observed, and were subject to, the outcomes of the choices that the partner had previously made. In reality, choices on social trials were generated using a standard decision model based on prospect theory, using parameters for a typical subject (see Methods). This procedure ensured that all participants had a similar experience in social trials.

When the partner chose the safe option, both players received the same outcome; when the partner chose the risky option, both players received the gamble. The critical manipulation centered on the independence of the two gambles for the subject and the partner. This meant that for any single gamble chosen by the partner, the outcomes experienced by the subject and their partner could be identical or different (Figure 2.1B), providing the potential for inequality. In all trials, the outcomes for the subject counted toward overall earnings. To investigate the relationship between the outcomes of choices and subjective emotional state, including choices made by others, we used experience sampling[15,27,28], repeatedly asking subjects, 'How happy are you at this moment?' after every 2-3 trials. Subjects were tested using two slightly different procedures (see Methods) with an identical trial structure, and were informed of total earnings only after completion of all tasks. Although happiness due to inequality could potentially be

measured without any choice on the subject's part, not being able to make any choices would reduce engagement, and the results of our previous studies show that outcomes resulting from a subject's choices substantially impact happiness. Thus, we interleaved social and non-social trials, and outcomes for the two types of trials were independent, allowing us to dissociate these influences.

2.4.1 Reinforcement learning model predicts subjective well-being

We first examined the determinants of subjective well-being, and found, consistent with our previous research[15], that subjects reported greater average happiness at the subsequent rating after winning compared to losing gambles in both social and non-social trials (Wilcoxon signed-rank test, n=47, both Z>4, P<0.001). In social trials, we tested whether there was an impact of partner outcomes on well-being by z-scoring ratings for each subject and computing average happiness at the subsequent rating across the following four contexts: both participants win, both participants lose, subject wins and partner loses, and subject loses and partner wins. The last two conditions are associated with advantageous (subject has more) and disadvantageous (subject has less) inequality, respectively. These two contexts are ones that might engender the social emotions of guilt and envy, respectively, emotions that might relate to terms in models of altruistic behavior[12].

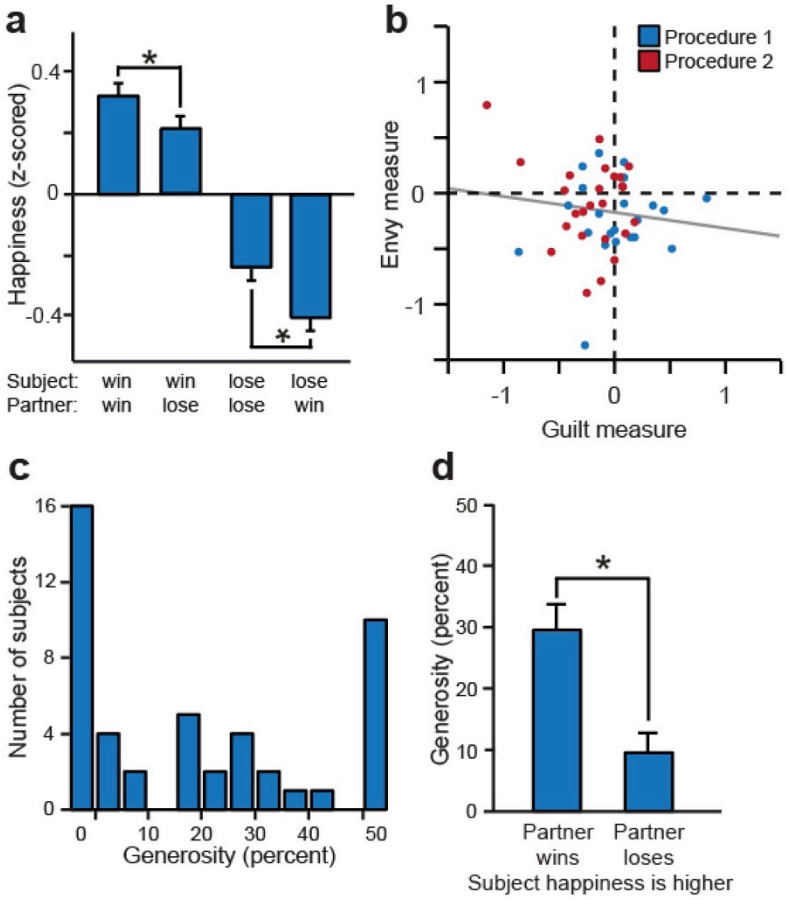


Figure 2.2 | Descriptive analysis (A) Subjects (*n*=47) reported being happier at the subsequent rating after winning compared to losing gambles, and happiness ratings were lower on average when the partner received a different outcome, regardless of whether that outcome was better or worse. **(B)** The amount that happiness was affected by advantageous inequality (guilt is when subject wins and is equal to happiness for partner loses minus partner wins) and disadvantageous inequality (envy is when subject loses and is equal to happiness for partner win minus partner loses) was uncorrelated across subjects (Spearman's ρ=-0.04, P=0.78). **(C)** Subjects completed a dictator game in which they could anonymously give a fraction of an endowment to their partner; 10 subjects gave half of the endowment and 16 subjects gave nothing. **(D)** Subjects were more generous in the dictator game if their happiness in the separate social decision task was higher when the partner won than lost gambles. The difference between guilt and envy measures was correlated with generosity in the dictator game (Spearman's ρ=-0.48, P<0.001). Subjects who were happier after the partner lost than won gambles included only 2 of 10 subjects who gave half but 12 of 16 subjects who gave nothing. Error bars, SEM. *P<0.05.

2.4.2 Inequality reduces subjective well-being

We found that, regardless of whether subjects themselves won or lost, average subjective well-being was attenuated by inequality in outcomes. Well-being was reduced both when subjects were better off (Z=-2.2, P=0.028) and when they were worse off (Z=-2.8, P=0.005) than the other person (Figure 2.2A). We tested whether the sensitivity of subjective well-being to advantageous and disadvantageous inequality (referred to here as guilt and envy) is equivalent, as might be expected if it reflected a unified concept of inequality aversion. However, we found no correlation between the change in well-being when subjects were better compared to worse off than their partner (Spearman's ρ=-0.04, P=0.78; Figure 2.2B), suggesting independent variation in the degrees of guilt and envy, a result inconsistent with a unitary concept of inequality aversion.

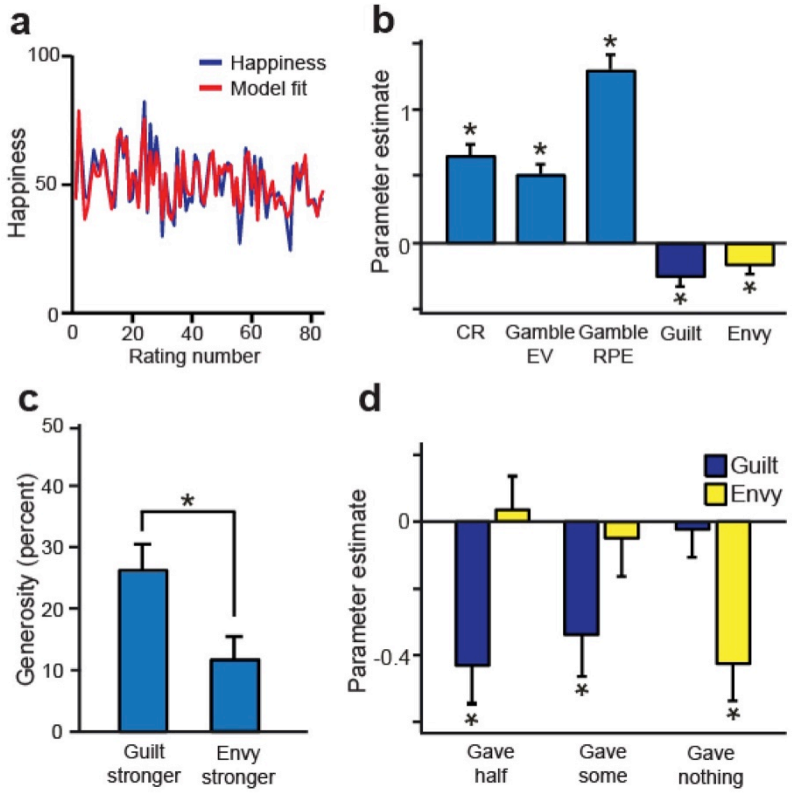


Figure 2.3 | Model-based analysis of happiness (A) Happiness ratings of an example subject over the course of the experiment plotted with the predictions of the guilt-envy model. **(B)** Happiness was affected by model parameters (n=47) related to the subject's rewards. Two additional model parameters related to inequality aversion were both negative, indicating that

both advantageous inequality (guilt parameter) and disadvantageous inequality (envy parameter) negatively impact happiness on average. **(C)** Subjects with stronger (more negative) guilt parameters were more generous in the separate dictator game than subjects with stronger (more negative) envy parameters. The difference between guilt and envy parameters was correlated with generosity in the dictator game (Spearman's ρ=-0.48, P<0.001). **(D)** Guilt and envy parameters estimated by the model for subjects with different levels of generosity in the dictator game. Subjects who gave nothing had significant envy parameters. Subjects who gave something had significant guilt parameters. Error bars, SEM. *P<0.05.

2.4.3 Emotional impacts of inequality relate to generosity

We next determined whether a social influence on well-being was related to generosity in the separate dictator game. For each subject, we computed the difference in happiness between the two situations where the partner loses and the two situations where the partner wins, equivalent to taking the difference between guilt and envy measures. Against a backdrop of a 20% average allocation in the dictator game (Figure 2.2C), consistent with previous studies[26,29], subjects who were on average happier after partner wins than partner losses were also more generous in the dictator game, compared to subjects who were less happy after partner wins than partner losses (Wilcoxon rank-sum test, Z=3.4, P<0.001; Figure 2.2D). Strikingly, the first group of subjects gave three times as much of an endowment on average than the second group (30% vs. 10%). Generosity in the dictator game was highly correlated with the difference between guilt and envy measures derived from happiness ratings (Spearman's ρ=-0.48, P<0.001). Although generosity might be thought to depend on the guilt of receiving an unexpected endowment, we found that generosity was not significantly correlated with guilt measures alone (Spearman's ρ=-0.18, P=0.22) but was correlated with envy measures (Spearman's ρ=0.30, P=0.042) such that subjects exhibiting greater envy were less generous, a pattern of results inconsistent with any plausible demand characteristics of the experimental design.

2.4.4 Modelling the impact of inequality aversion on well-being

Our next goal was to apply our previously established methodology for measuring determinants of momentary subjective well-being to quantify individual dispositions in the social domain that impact emotional reactivity. Our starting point was our pre-existing non-social happiness model[15], in which chosen certain rewards (CR), the expected value (EV) of chosen gambles, and

RPEs resulting from those expectations, all exert separate influences that decay exponentially with time:

$$\mathbf{Happiness}(t) = w_0 + w_1 \sum_{j=1}^{t} \gamma^{t-j}\, \mathbf{CR}_j + w_2 \sum_{j=1}^{t} \gamma^{t-j}\, \mathbf{EV}_j + w_3 \sum_{j=1}^{t} \gamma^{\mathrm{t-j}}\, \mathbf{RPE}_j$$

Equation 2.1

where t is trial number, w_0 is a constant term, other weights w capture the influence of different event types, $0 \le \gamma \le 1$ is a discount factor that makes events in more recent trials more influential than those in earlier trials, $\mathbf{CR}_j$ is the certain reward if chosen instead of a gamble on trial j, $\mathbf{EV}_j$ is the average reward for the gamble if chosen on trial j, and $\mathbf{RPE}_j$ is the RPE on trial j contingent on choice of the gamble. Terms for unchosen options were set to zero. We z-scored happiness ratings to prevent subjects with greater variability in their ratings from having a disproportionate effect on results. The constant term is omitted when ratings are z-scored. We fitted parameters to the happiness ratings of individual subjects in the social decision task and found, as expected, that **CR**, **EV**, and **RPE** weights were on average positive (all $Z>4$, $P<0.001$). The discount factor γ was 0.67±0.25 (mean±SD) indicating that ratings on average depended on the cumulative impact of 5-10 past events. Despite having no way to account for any social effect, this model explained momentary happiness well, with r^2=0.39±0.19 (mean±SD), comparable to fits for a non-social task in a previous study[15].

We next expanded the model by including additional terms to account for influences related to advantageous and disadvantageous inequality[12]. These influences might be considered as related to guilt and envy, respectively:

$$\mathbf{Happiness}(t) = w_0 + w_1 \sum_{j=1}^{t} \gamma^{t-j}\, \mathbf{CR}_j + w_2 \sum_{j=1}^{t} \gamma^{t-j}\, \mathbf{EV}_j + w_3 \sum_{j=1}^{t} \gamma^{\mathrm{t-j}}\, \mathbf{RPE}_j$$

$$+ w_4 \sum_{j=1}^{t} \gamma^{\mathrm{t-j}} \max\left(\mathbf{R}_j - \mathbf{O}_j, 0\right) + w_5 \sum_{j=1}^{t} \gamma^{t-j} \max\left(\mathbf{O}_j - \mathbf{R}_j, 0\right)$$

Equation 2.2

where w_4 relates to advantageous inequality (guilt) when the reward received by the subject $\mathbf{R}_j$ exceeds the reward received by the other player $\mathbf{O}_j$, and w_5 relates to disadvantageous inequality (envy) when $\mathbf{O}_j$ exceeds $\mathbf{R}_j$. This guilt-envy model explained momentary happiness better than its non-social variant with r^2=0.44±0.18 (mean±SD; Figure 2.3A). This model was preferred to the simpler non-social model in a Bayesian model comparison[30,31] (see Table 2.1), demonstrating that social comparison significantly impacts subjective well-being in our task. Model parameters for guilt and envy were negative on average (both Z<-2, P<0.05; Figure 2.3B), consistent with both advantageous and disadvantageous inequality reducing momentary happiness.

2.4.5 Envy and guilt parameters predict generosity

When we tested how model parameters related to individual social preferences, we found subjects with stronger (more negative) guilt parameters were more generous in the dictator game than subjects with stronger (more negative) envy parameters (Z=2.8, P=0.006; Figure 2.3C). Consistent with the descriptive analysis, the difference between guilt and envy parameters estimated from happiness ratings was highly correlated with generosity in the dictator game (Spearman's ρ=-0.48, P<0.001). Guilt but not envy parameters were significantly negative for subjects that altruistically gave either half or some of the endowment (guilt, both Z<-2.3, P<0.05; envy, both $|Z|$<0.5, P>0.5; Figure 2.3D) while the opposite was true for those subjects that gave nothing (guilt, Z=-0.2, P=0.88; envy, Z=-2.9, P=0.003).

One concern is whether demand characteristics might contribute to any of our results. Some subjects might have noticed that inequality was one feature of the experiment and hypothesized that well-being should reflect a unitary concept of inequality ("inequality is bad"). To test whether this possibility could explain our results, we fitted an additional model with a term for the magnitude of the difference in rewards between players. The inequality parameter in this simple-inequality model was significantly negative on average (Z=-5.13, P<0.001), capturing lower well-being with greater inequality (see Methods). However, this inequality parameter was uncorrelated with generosity in the dictator game (Spearman's ρ=-0.036, P=0.81), which might theoretically have responded to the same demand characteristic, and the guilt-envy model outperformed the simple-inequality model according to Bayesian model comparison (Table 2.1).

Model	Parameters per subject	Mean r^2	Median r^2	Model BIC	Relative BIC
Non-social	4	0.39	0.37	-1723	+72
Simple-inequality	5	0.41	0.40	-1704	+91
Guilt-envy	6	0.44	0.47	-1795	0

Table 2.1| Bayesian model comparison analysis

BIC values are summed across the 47 subjects. Model fits were performed with z-scored happiness ratings. All three models contained separate terms for CRs, gamble EVs, and gamble RPEs, with influences that decay exponentially. The simple-inequality model included an additional parameter for the magnitude of the difference in outcomes between the two players. The guilt-envy model included additional parameters for advantageous and disadvantageous inequality. The guilt-envy model had the lowest BIC and is therefore preferred by Bayesian model comparison.

2.5 Discussion

Our results provide striking quantitative confirmation that an individual's subjective reports of momentary well-being reflect not just how well things are going relative to expectations, but also how things are going relative to other people, even when outcomes for others are both independent from, and irrelevant to, the subjects' own earnings. By quantifying social preferences based on emotional reactivity to inequality separately from economic choice, we avoid strategic considerations, a potential confound in using economic games to understand the role of inequality aversion in behavior[32]. Furthermore, by using the continuously varying subjective state as an output measure, we avoid forcing subjects to explicitly admit to emotions that might be perceived to have negative social connotations, such as envy.

Increasing wealth disparity in many countries[33,34] lends urgency to the need to understand the impact of inequality, both at individual and societal levels[35]. Our demonstration that inequality aversion reduces momentary well-being aligns with wider observations of inequality's negative impact on societal well-being[36]. Culturally entrained aversion to inequality such as 'Janteloven' observed in Scandinavia[37] may therefore play an important role in shaping well-being in those

countries, which rank highly in international well-being surveys. The fact that individual differences in well-being measures were predictive of social preferences suggests these parameters reflect values that are at least part of a stable variation in generosity between individuals[38], variation that has been difficult to explain using economic approaches alone[14]. Our findings also highlight an important issue for future research: our subjects experienced inequality in social trials where they were unable to influence their partner's decision. Understanding how instrumental control impacts well-being could shed additional light on the mediation of societal inequality.

We adopted a quantitative model that opens up new avenues to investigate the relationship between subjective well-being and behavior. Although numerous studies have linked experienced and anticipated emotions during choice to subsequent behavior (reviewed in ref. 39), it has remained unclear whether choices accurately anticipate the emotional impact of outcomes on subjective well-being, with some arguing that these quantities are distinct[40]. However, recent economic research suggests quantitative links can be forged between hypothetical choices and hypothetical consequences for well-being[41]. Here, we demonstrate a precise link between subjective well-being following actual rewards and incentivized economic altruistic choice.

There is considerable debate as to the underlying basis for altruistic behavior. Although Fehr and Schmidt posit that guilt and envy relate to altruistic behavior[12], it has never been tested whether emotional responses to advantageous and disadvantageous inequality explain a heterogeneity in generosity[14]. We found that the relative emotional impact of guilt and envy is predictive of generosity, lending support to the Fehr-Schmidt model. However, our results are inconsistent with two assumptions of this model, specifically that weights for guilt and envy are correlated and that the weight for envy is greater than the weight for guilt. We found that the emotional impacts of advantageous and disadvantageous inequality are uncorrelated such that people who experience more guilt do not necessarily experience more envy. Furthermore, we found that guilt parameters were greater than envy parameters in the majority of our subjects, in sharp contrast to a stated assumption of the Fehr-Schmidt model. This result is relevant to altruistic behavior in that participants who had larger guilt than envy parameters were on average more generous that individuals for whom the converse was true (Figure 2.3C). Our

results therefore recommend alternate social-welfare preference models that relax the assumptions of the Fehr-Schmidt model[42,43].

Our emotional dissection of inequality aversion also addresses an important critique that emerges from the constraints of dictator games, namely that any action other than keeping all of the money looks like inequality aversion[42]. Inattentive subjects could inadvertently appear altruistic. Our results show that much of the variance in generosity cannot be explained by this concern, because noisy happiness ratings could not be misconstrued as evidence for inequality aversion. Our finding of a link between emotional measures and generosity provide a new perspective on the value of dictator games as assays of social preferences.

Recent work on the ontogeny of fairness across cultures finds that an aversion to disadvantageous inequality arises early in development, but that an aversion to advantageous inequality arises later, and possibly for strategic reasons[13]. Similarly, dual-process models suggest that prosocial behavior might result from an interaction between intuitive/emotional and deliberative/non-emotional processes[44,45]. However, we find that emotional processes alone are sufficient to explain heterogeneity in generosity, without invoking strategic concerns or conflict between emotional and non-emotional processes. This result also argues against the need to appeal to any understanding of the emotional state of another person, as in popular empathy models[46], at the time that altruistic decisions are made.

Demand characteristics can be a concern in the study of both well-being and altruism. However, there are several reasons why a demand-driven explanation of our results is unlikely. First, the task was designed such that the rewards of others are irrelevant to earnings, reducing the chance that subjects will realize that their ratings reveal a socially undesirable emotion like envy. Evidence that this was successful is that most (~90%) of the variance in ratings accounted for by the model arose from non-social influences, influences known to be the same in paid lab subjects and unpaid anonymous subjects[15], inconsistent with significant demand effects in the non-social influences on well-being. Second, the average forgetting factor is such that ratings reflected the cumulative influence of 5-10 past events. However, ratings were made too quickly (on average in 1.7 s in the social decision task) to allow the sort of deliberate calculation that demand effects over such a timescale might require. Third, we fitted a simple-inequality model

that captures the unitary concept of inequality ("inequality is bad") that is most likely to be consistent with perceived experimenter demands. This model did not explain variation in generosity, and did not explain the data as well as the guilt-envy model. A final argument against any explanation in terms of demand characteristics is that we observe a strong effect of envy on well-being in subjects who also appear sufficiently immune to experimenter demands as to give nothing in the dictator game (Figure 2.3D).

Our computational approach might be fruitfully employed to meet a variety of challenges. The most immediate application is in testing hypotheses regarding the role of emotions in prosocial behavior across economic games, including social and moral emotions that might relate to behaviors such as trust and punishment. Furthermore, computational models such as ours can provide precise subject-specific predictions for interrogating the neural circuits that support prosocial behavior while also generating predictors related to negative emotions such as guilt and envy that can be difficult to elicit explicitly in experimental settings. Understanding individual differences in the determinants of well-being may also yield insight into interactions between people of different socioeconomic status, which may have economic implications. Finally, individual phenotyping based on emotional dynamics could provide a powerful tool to dissect social pathologies, such as borderline personality disorder.

2.6 References

1. Oswald, A. J. & Wu, S. Objective confirmation of subjective measures of human well-being: evidence from the U.S.A. *Science* **327**, 576-579 (2010).
2. Krueger, A. B. & Stone, A. A. Progress in measuring subjective well-being. *Science* **346**, 42-43 (2014).
3. Stiglitz, J. E., Sen, A. & Fitoussi, J. P. Report of the Commission on the Measurement of Economic Performance and Social Progress (2009).
4. Blanchflower, D. G. & Oswald, A. J. Well-being over time in Britain and the USA. *J. Pub. Econ.* **88**, 1359-1386 (2004).
5. Layard, R. Happiness: Lessons from a New Science (Allen Lane, 2005).
6. Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O. & Zweig, J. S. The happiness-income paradox revisited. *Proceedings of the National Academy of Sciences* **107**, 22463-22468 (2010).
7. Kahneman, D. & Deaton, A. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences* **107**, 16489-16493 (2010).
8. Stevenson, B. & Wolfers, J. Subjective well-being and income: is there any evidence of satiation? *Am. Econ. Rev.: Papers Proc.* **103**, 598-604 (2013).
9. Clark, A. E. & Oswald, A. J. Satisfaction and comparison income. *J. Pub. Econ.* **61**, 359-381 (1996).
10. Luttmer, E. F. P. Neighbors as negatives: relative earnings and well-being. *Q. J. Econ.* **120**, 963-1002 (2005).

11. Boyce, C. J., Brown, G. D. A. & Moore, S. C. Money and happiness: rank of income, not income, affects life satisfaction. *Psychol. Sci.* **21**, 471-475 (2010).
12. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817-868 (1999).
13. Blake, P. R. *et al.* The ontogeny of fairness in seven societies. *Nature* **528**, 258-261 (2015).
14. Blanco, M., Engelmann, D. & Normann, H. T. A within-subject analysis of other-regarding preferences. *Games Econ. Behav.* **72**, 321-338 (2011).
15. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences* **111**, 12252-12257 (2014).
16. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593-1599 (1997).
17. Caplin, A., Dean, M., Glimcher, P. W. & Rutledge, R. B. Measuring beliefs and rewards: a neuroeconomic approach. *Q. J. Econ.* **125**, 923-960 (2010).
18. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. Dopaminergic modulation of decision making and subjective well-being. *J. Neurosci.* **35,** 9811-9822 (2015).
19. Fliessbach, K. *et al.* Social comparison affects reward-related brain activity in the human ventral striatum. *Science* **318**, 1305-1308 (2007).
20. Tricomi, E., Rangel, A., Camerer, C. F. & O'Doherty, J. P. Neural evidence for inequality-averse social preferences. *Nature* **463**, 1089-1091 (2010).
21. Zaki, J. & Mitchell, J. P. Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences* **108**, 19761-19766 (2011).
22. Sokol-Hessner, P. *et al.* Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences* **106**, 5035-5040 (2009).
23. Brown, H. R. *et al.* Crowdsourcing for cognitive science - the utility of smartphones. *PLoS ONE* **9**, e100662 (2014).
24. Kahneman, D., Knetsch, J. L. & Thaler, R. H. Fairness and the assumptions of economics. *J. Business* **59**, S285-S300 (1986).
25. Eckel, C. C. & Grossman, P. J. Altruism in anonymous dictator games. *Games Econ. Behav.* **16,** 181-191 (1996).
26. Camerer, C. Behavioral Game Theory: Experiments in Strategic Interaction (Princeton University Press, 2003).
27. Csikszentmihalyi, M. & Larsen, R. Validity and reliability of the experience sampling method. *J. Nerv. Ment. Disease* **175**, 526-537 (1987).
28. Killingsworth, M. A. & Gilbert, D. T. A wandering mind is an unhappy mind. *Science* **330**, 932 (2010).
29. Engel, E. Dictator games: a meta study. *Exp. Econ.* **14**, 583-610 (2011).
30. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461-464 (1978).
31. Burnham, K. P. & Anderson, D. R. *Model Selection and Inference* (Springer, 1998).
32. Fehr, E. & Krajbich, I. Social preferences and the brain. *Neuroeconomics: Decision-Making and the Brain, 2nd Ed.* (eds Glimcher, P. W. & Fehr, E.) 193-218 (Academic Press, 2014).
33. Piketty, T. *Capital in the Twenty-First Century* (Belknap Press, 2014).
34. Piketty, T. & Saez, E. Inequality in the long run. *Science* **344**, 838-843 (2014).
35. Nishi, A., Shirado, H., Rand, D. G. & Christakis, N. A. Inequality and visibility of wealth in experimental social networks. *Nature* **526**, 426-429 (2015).
36. Wilkinson, R. G. & Pickett, K. E. Income inequality and social dysfunction. *Ann. Rev. Sociol.* **34**, 493-511 (2009).
37. Nelson, M. R. & Shavitt, S. Horizontal and vertical individualism and achievement values: a multimethod examination of Denmark and the United States. *J. Cross-Cultural Psychol.* **33**, 439-458 (2002).
38. Fisman, R., Kariv, S. & Markovits, D. Individual preferences for giving. *Am. Econ. Rev.* **97**, 1858-1876 (2007).

39. DeWall, C. N., Baumeister, R. F., Chester, D. S. & Bushman, B. J. How often does currently felt emotion predict social behavior and judgment? A meta-analytic test of two theories. *Emotion Rev.* 1-8 (2015).
40. Kahneman, D., Wakker, P. P. & Sarin, R. Back to Bentham? Explorations of experienced utility. *Q. J. Econ.* **112**, 375-406 (1997).
41. Benjamin, D. J., Heffetz, O., Kimball, M. S. & Szembrot, N. Beyond happiness and satisfaction: toward well-being indices based on a stated preference. *Am. Econ. Rev.* **104**, 2698-2735 (2014).
42. Charness, G. & Rabin, M. Understanding social preferences with simple tests. *Q. J. Econ.* **117**, 817-869 (2002).
43. Engelmann, D. How not to extend models of inequality aversion. *J. Econ. Behav. Org.* **81**, 599-605 (2012).
44. Rand, D. G., Greene, J. D. & Nowak, M. A. Spontaneous giving and calculated greed. *Nature* **489**, 427-420 (2012).
45. Rand, D. G. *et al.* Social heuristics shape intuitive cooperation. *Nat. Commun.* **5**, 3677 (2014).
46. Batson, C. D. & Shaw, L. L. Evidence for altruism: toward a pluralism of prosocial motives. Psychol. Inquiry **2**, 107-122 (1991).

Chapter 3: Acute stress selectively impairs learning to act

Previously published as:

de Berker A. O.*, Tirole M.*, Rutledge R. B., Cross G. F., Dolan R. J. & Bestmann S. Acute stress selectively impairs learning to act. *Scientific Reports* **6,** 29816 (2016).

* Denotes shared first-authorship

3.1 Abstract

The experiments reported in **Chapter 2** revealed that emotional dynamics can be predicted using models inspired by reinforcement learning. In this experiment we examine the complementary relationship; how does an intense emotional experience alter subsequent reward learning? Previous work has established that stress interferes with instrumental learning. However, choice is also influenced by non-instrumental factors, most strikingly by biases arising from Pavlovian associations that facilitate action in pursuit of rewards and inaction in the face of punishment. This suggests, contrary to classical models from reinforcement learning, that action and valence play interactive and partially separable roles in animal learning. Whether stress impacts on instrumental learning via these Pavlovian associations is unknown. Here, in a task where valence (reward or punishment) and action (go or no-go) were orthogonalised, we asked whether the impact of stress on learning was action or valence specific. We exposed 60 participants either to stress (socially-evaluated cold pressor test) or a control condition (room temperature water). We found that stress specifically impaired learning to produce an action, irrespective of the valence of the outcome, an effect consistent with a Pavlovian linkage between punishment and inaction. This deficit in action-learning was also reflected in pupillary responses; stressed individuals showed attenuated pupillary responses to action, hinting at a noradrenergic contribution to impaired action-learning under stress.

3.2 Introduction

Stress is linked to a cascade of changes in central and peripheral physiology, including a rapid rise in catecholamines and a sustained increase in glucocorticoids [1]. Stress also induces a shift in instrumental learning from flexible (model-based) systems towards an inflexible, and experience-dependent (model-free), form of control [2-6]. The degree of shift depends upon working memory capacity [5], consistent with the idea that an impaired allocation of cognitive resources is a key mediating mechanism [2-7].

However, recent work has also highlighted the importance of non-instrumental influences upon choice, as reviewed in section 1.3.2.2. Pavlovian responses resulting from the prediction of reward and punishment interfere with the learning of instrumental contingencies [8-11]. For example, this innate coupling between reward and approach is of such potency that chicks are unable to learn to walk away from a food source in order to harvest it [12]. An opposite linkage, between punishment and inaction, is evident in the freezing behaviour (inaction) seen across a wide variety of species [13].

Pavlovian approach behaviour is also thought to underlie aspects of cue-triggered relapse in addiction [14], an influence potentiated by stress [15] via putative corticosteroid-dopaminergic interactions [16]. The observation that stress also engenders fast, context-inflexible, forms of control [2], raises the possibility that its impact on instrumental behaviour is mediated via Pavlovian biases. Limited evidence in rodents [17] suggests that stress transiently attenuates the impact of reward prediction upon vigour (Pavlovian-instrumental transfer, PIT), although other studies failed to observe an effect [18]. However, stress itself constitutes a sustained aversive state. Given the linkage between negative valence and inaction [11], this suggests an alternative hypothesis, namely that stress impacts on instrumental learning via the Pavlovian coupling of punishment and inaction. Such a coupling would explain the augmented inhibition of pre-potent actions observed under threat of shock [19], and predicts that sustained activation of punishment-related tendencies (such as active avoidance, or inaction[11]) might interfere with the learning of instrumental responses which required different behavioural outputs (such as approach).

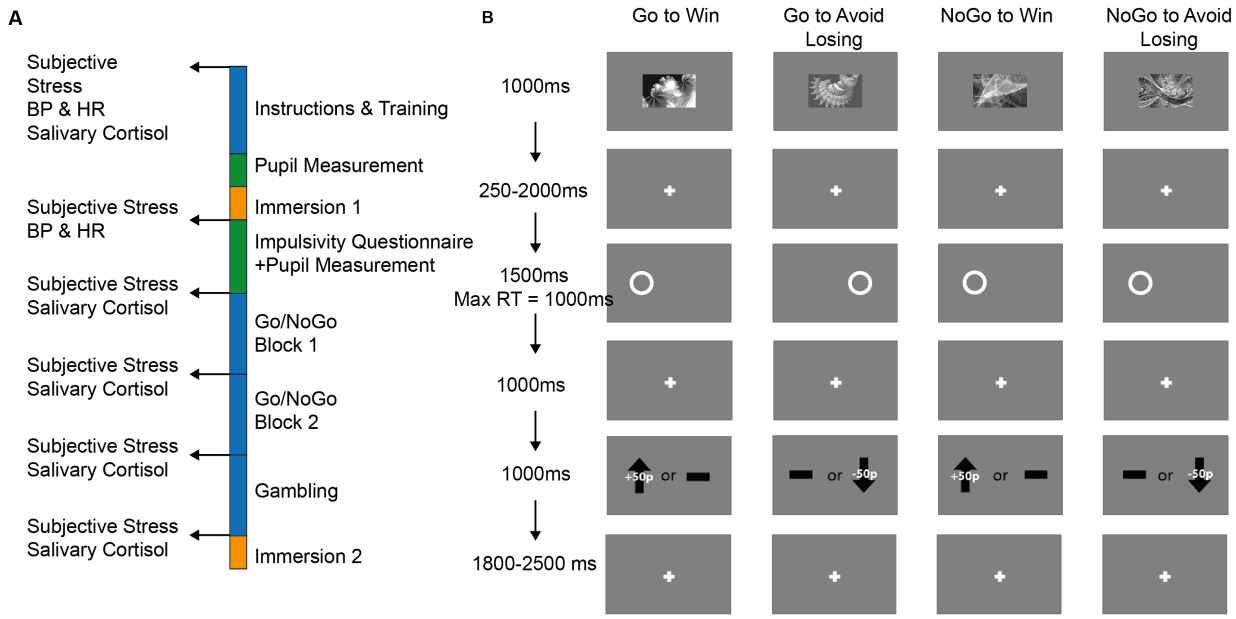


Figure 3.1 | Experimental design (A) Timecourse of the experiment. After instruction and training, subjects were asked to immerse their arms for three minutes in 0-1 ^{0}C (Stressed group) or 24-27 ^{0}C (Control group) water. Blood Pressure (BP) and Heart Rate (HR) were assessed before and after immersion. After a waiting period of ten minutes, they then completed the Go/NoGo task followed by a separate gambling task. At the end of the experiment, subjects underwent an expected second immersion lasting 30 seconds. Subjective stress and salivary cortisol were assayed regularly throughout the task. **(B)** The Go/NoGo task. On each trial, one of four cues was presented. Each cue was associated with an initially unknown correct action (Go or NoGo) and an outcome (Win money or Avoid Losing money). Subjects learned to select actions based upon the outcomes. Following the cue, a target was presented on the left or right side of the screen and subjects chose whether to press a button corresponding to the side of target presentation. For the cues associated with winning money, the correct choice was rewarded with an increase in earnings on 80% of trials. For cues associated with losing money, the incorrect choice was punished with a decrease in earnings on 80% of trials.

To distinguish between these two competing hypotheses we compared a Stressed group who underwent the socially-evaluated Cold Pressor Test (CPT), and a Control group who submerged their hands in room-temperature water [5-7] (Figure 3.1A). We measured subjective stress and salivary cortisol at multiple time-points, allowing us to track stress levels throughout the experiment. Additionally, we recorded pupil diameter throughout, as an indirect assay of noradrenergic activity [20,21], which has a central role both in action initiation [22] and the co-ordination of stress responses [23].

On each trial, participants saw a cue, and had to produce or withhold an action (Go/NoGo) so as to gain a reward (Win) or avoid a punishment (Avoid Loss) (Figure 1B). The design was factorial, giving four conditions: Go to Win, Go to Avoid Losing, NoGo to Win, NoGo to Avoid Losing. The contingencies were probabilistic, such that the correct action led to the better outcome in 80% of cases. This task can be understood as comprising two Pavlovian-congruent conditions, in which the instrumental requirements match the valence-evoked Pavlovian resposnes (Go to Win and NoGo to Avoid Loss), and two Pavlovian-incongruent conditions, where the instrumental and Pavlovian responses are in conflict (NoGo to Win and Go to Avoid Losing). This task therefore provides a well-validated assay of Pavlovian biases, assessed by comparison of performance in the Pavlovian-congruent conditions, on which people perform well, and Pavlovian-incongruent conditions, on which they perform poorly [8,9]. This enabled us to ask whether stress affected Pavlovian biases over learning, or whether its effects were better described by a specific bias towards inaction, as suggested by an inhibition account of stress effects upon behavior.

3.3 Methods

3.3.1 Participants

We recruited 64 participants (32 males) through UCL's Institute of Cognitive Neuroscience (ICN) database. Participants were screened for medical conditions, previous CPT exposure, and previous participation in any experiments involving the Go/NoGo task. 4 participants were excluded (distractibility during the experiment or misunderstanding of the task revealed upon debrief), leaving 60 participants (30 CPT, 30 Control, 15 males in each group).

All participants signed an informed consent form. All collected data was treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998. Medical supervision was present throughout. The protocol was approved by the UCL Research Ethics Committee (Ethics 4377/001). All experiments were conducted in accordance with approved guidelines.

3.3.2 Experimental Procedure

The experimental procedure is summarised in Figure 3.1A. Participants read an instruction sheet describing the structure of the experiment, which specified whether they were in the CPT or Control group. They then gave their informed consent, and underwent basic computerized

training in the Go/NoGo task and a separate gambling task (results not discussed here). They then provided a saliva sample and had their blood pressure taken, before a measurement of baseline pupil diameter. Participants then submerged an arm in ice-cold (Stressed condition) or room-temperature water (Control condition) [details below]. Following withdrawal of the hand from water, participants completed a questionnaire (Urgency, Premeditation, Perseverance, Sensation seeking; UPPS [26]), and then maintained fixation until 10 minutes had elapsed since the end of submersion. This period was chosen to accommodate the timecourse of glucocorticoid release, such that cortisol concentrations would be elevated when they began the Go/NoGo task [56]. Participants then performed the Go/NoGo task, with a self-paced break halfway through. They then performed a gambling task, the results of which will be presented elsewhere, before performing a final, brief submersion of 30 seconds.

3.3.3 Stress manipulation & measures

3.3.3.1 Cold Pressor Test

We used the widely adopted Cold Pressor Test (CPT) to induce stress [5-7,57,58]. Water temperature was 0-1^{0}C in the Stressed condition, and 24-27^{0}C in the Control condition. Participants were asked to keep their arm submerged for 3 minutes. Participants in the Stressed condition were monitored by an additional experimenter, who entered the room specifically to observe this phase of the experiment, adding a social-evaluation component to the stressor [7]. All control subjects kept their arms submerged for the full 3 minutes, whilst several stressed participants were unable to remain submerged for the entire period (mean duration=144 seconds, range 39–180 seconds). Subjects were informed at this point that they would be completing a second submersion at the end of the experiment, a manipulation designed to sustain negative affect throughout the task.

3.3.3.2 Cortisol

We measured stress responses in three ways, designed to capture the subjective response (assessed with visual analogue scale ratings) and physiological responses to stress. The latter can be fractionated into a rapid, catecholaminergic component (indexed by pupil diameter) and a slower, glucocorticoid release (indexed by salivary cortisol samples). We refer to the timings of cortisol and subjective stress assays with reference to the end of submersion and the approximate times for subsequent assays: T0, T10, T20, T30, T45. Cortisol was not measured at

T0, as the sluggish dynamics of the HPA axis preclude an immediate glucocorticoid response to stress [59]. Pupil diameter was measured throughout, though measurements could not be obtained at hand immersion and withdrawal due to excessive movement.

3.3.3.3 Subjective stress

Subjective stress was assayed using experiential sampling [43,60], which involved each subject moving a cursor along a line to answer the question 'How stressed do you feel right now?', anchored by 'Not stressed' and 'Very stressed' at each extreme. Ratings for subjective stress from three participants were lost due to technical error.

Pupil diameter was measured using an EyeLink 1000 system, sampling at 250Hz. The experiment took place in a darkened room, with a computer screen shielded on each side to minimise reflections. Participants were asked to maintain fixation wherever possible. All stimuli in the experiment were luminance matched within stimulus-type, although it was not possible to match luminance between stimuli. Data was subsequently exported to ASCII and imported to Matlab for analysis, where automatically-identified blink events were removed and replaced via linear interpolation of samples 140ms either side of the blink. Data were then low-pass filtered (2^{nd} order Butterworth Filter, 4Hz) and z-scored before analysis [31,32,61].

3.3.3.4 Pupil diameter

For analyses of stress effects upon pupil diameter (Figure 3.2C) we measured pupil diameter in a baseline period prior to T0, during immersion, and post immersion. A single subject was excluded from this analysis due to complete data loss during immersion. Having removed periods of signal dropout, we then subtracted the average pupil diameter for each subject during the baseline period to correct for inter-individual variance in pupil size.

To examine the effect of task-events upon pupil diameter, we epoched the data in three ways: by Cue, by Target, and by Outcome. This was necessary as the timings of events within each trial varied according to imposed jitter and variable reaction time (Figure 3.1B). In all cases, we accounted for drift in baseline pupil diameter over time by subtracting a baseline measurement of the average diameter for the first 200ms following cue presentation, resulting in pupil diameter traces for all trials starting around zero.

We used multiple regression models to decompose the influence of different task events upon pupil diameter at multiple timepoints [30]. We used separate models in order to describe pupil diameter locked to Target presentation, and at Outcome. In both, we assessed the influence of action (Go/NoGo), outcome valence (Gain, No Change, Loss) and surprise (absolute prediction error from reinforcement learning models, see below) upon pupil responses. In the Outcome analysis, we also included an interaction term (Action*Valence). All predictors were z-scored before analysis to produce regression coefficients of comparable magnitude.

We decomposed each epoch into 40 time points (one every 200ms), and performed a multiple regression analysis across trials at each time point using the robustfit function in Matlab. This allowed us to examine the influence of each of our predictors (action, valence, and surprise) upon pupil diameter at different times in the trial. Inference was based upon the distribution of regression coefficients (β) across subjects, where we test for differences from zero using single-sample t-tests. When testing for differences from zero (i.e. whether a predictor consistently affects pupil diameter) we take the average β coefficients for a period 1-2s following outcome. For inference between individuals, we use the maximum absolute β across timepoints for each subject, allowing for interindividual variance in reaction time and pupillary dynamics.

3.3.3.5 Cortisol analysis

For salivary samples, participants salivated through straws into 2ml polypropylene tubes. Samples were frozen on the day of collection. Analysis was performed by Viapath at King's College Hospital, using a competitive immunoassay. Briefly, cortisol in the sample competes with cortisol conjugated to horseradish peroxidase for binding sites on a microtitre plate. Unbound reagents are then washed away. Bound cortisol enzyme conjugate is measured by the reaction of the horseradish peroxidase enzyme to the substrate tetramethylbenzidine, producing a blue colour. A yellow colour is formed after stopping the reaction with an acidic solution. The concentration of cortisol in the sample is calculated as a function of the optical absorption at 450 nm; more absorption implies greater concentration of cortisol enzyme conjugate, and therefore lower concentration of cortisol in the sample (for further details see [62]). We quantified the experimental manipulation-evoked change in cortisol by calculating the Area Under Curve (AUC) with respect to baseline cortisol concentrations [29]. This standardised measure is a discrete equivalent of integrating the timecourse of cortisol concentrations.

3.3.4 Go/NoGo task

The structure of the Go/NoGo task [9,25,33,37] is depicted in Figure 3.1B. Participants performed 240 trials. On each trial, one of four stimuli was displayed. Each stimuli denoted a different action/outcome contingency: Go to Win, NoGo to Win, Go to Avoid Losing, and NoGo to Avoid Losing. Following a variable interval, a target was presented on the left or right hand side of the screen, to which participants decided to respond (pressing the arrow button corresponding to the side of target presentation) or not. Following a brief fixation, the outcome was then presented. In the Win conditions, the correct choice was associated with a monetary gain on 80% of trials, and no change in earnings on 20%. In Avoid Losing trials, correct choice resulted in no change in earnings in 80% of trials, and a monetary loss in 20% of trials. Following incorrect choice, the outcomes were flipped, such that the worst outcome was received 80% of the time. Trials were separated with a variable fixation period of 1800-2500ms.

We analysed data from the task in four ways. Firstly, we examined average learning curves for each condition and in each group, following previous analyses [9]. In analyses where we present aggregates across conditions (e.g. Fig 3.3B and 3.3C), we averaged across conditions (within subject) before computing between-subject averages. Where relevant we performed two-sample t-tests between groups, correcting for comparisons at multiple time-points using the Benjamini-Hochberg method [63] to control the False Discovery Rate (FDR).

However, average learning curves can obscure more discrete differences between individuals who have learnt the task and those who have not [28]. To accommodate this, we also performed group comparisons based upon the number of participants who responded correctly in >50% of trials for each condition, testing for significance using chi-squared tests.

Thirdly, modified reinforcement learning models describe learning in the Go/NoGo task [11] as a process of belief updating in each state on the basis of reward prediction errors, where the latter capture discrepancies between an action's actual and expected consequences. To test for effects of stress upon the Pavlovian bias we fit a common variant of Q-learning [34] in which action-values are biased by an interaction between action and state-value, amplifying weights on Go responses in the reward domain and suppressing them in the punishment domain (see [11], Box 1 for a limpid description of this implementation, and below). We also use a representation

of reward prediction error in this model to quantify and control for effects of surprise upon pupil diameter [30,31].

Finally, the Pavlovian performance bias was calculated as described previously [37] as the average of action invigoration in rewarded conditions ([Go in Go To Win + Go in NoGo to Win]/Total Go) and action suppression in punished conditions ([NoGo in Go to Avoid Losing + NoGo in NoGo to Avoid Losing]/Total NoGo). This provided a summary measure of how strongly action and valence interacted in choice.

3.3.5 Structure of the Pavlovian bias model

Our model is based upon learning in 4 states – Go to Win, NoGo to Win, Go to Avoid Losing, NoGo to Avoid Losing. These correspond to the 4 cues. For each state, there are two possible actions: Go and NoGo. We assume that on each trial the agent updates their beliefs about which action is betterbased upon Reward Prediction Errors (RPEs), which are calculated as the difference between outcomes and the agent's expectation. Expectations are captured by Q-values (hence Q-learning). See section 1.3.2.1 for a full description of Q-learning and its antecedents. Here, $Q_t(a,s) +$ describes the Q-value of state *s* and action *a,* at time *t*. Q-values are updated via RPE's, so following an action *a* in state *s:*

$$Q_{t+1}(a,s) = Q_t(a,s) + \alpha[\rho r_t - Q_t(a,s)]$$

Equation 3.1

Where Q_t is the state value on trial t, α is the learning rate, ρ is a weighting parameter which is multiplied by r_t, the reward or punishment on that trial (taking a value of -1, 0 (no change) or 1). We allow ρ to vary, accommodating varying sensitivity to punishments and rewards (losing a given amount of money is typically more negative than acquiring the same amount of money). To reiterate, *a* and s denote the current action and the current state, respectively.

Action selection is based upon $Q_t(a,s)$; however, the model is 'indirect' in the sense that action weight are allowed to diverge from $Q_t(a,s)$; by the addition of two additional terms: an action bias, and a Pavlovian bias. This captures the intuition that action selection might be subject to a variety of biases that do not affect beliefs. Action weights are thus updated separately from Q_t, following Q-updates, and contingent upon which action was selected:

If Go: $W_{t+1}(a,s) = Q_{t+1}(a,s) + \varepsilon + \pi V_t(s)$

If NoGo: $W_{t+1}(a,s) = Q_{t+1}(a,s)$

Equation 3.2

Where ε is the action bias, π is the Pavlovian bias, and $V_t(s)$ is the current state value. The effect of this uncoupling between *Q* and *W* is to accommodate an action bias (a general bias towards Go), and the Pavlovian interaction between action and valence, which emerges from the parameter π as follows. Go weights in rewarded states are amplified by the addition of the product of the positive parameter π and the positive state value $V_t(s)$. Conversely, negative state values decrease W_{t+1} for Go trials, through the addition of the product of the positive parameter π and the negative state value $V_t(s)$. Note that unlike Q_t, V_t is a function of the state alone, not contingent upon the currently selected action. V_t is updated through prediction errors in a manner analogous to Q_t but irrespective of action:

$$V_{t+1}(s) = V_t(s) + \alpha[\rho r_t - V_t(s)]$$

Equation 3.3

Where $V_t(s)$ is the state value of state *s* at time *t*, α is the learning rate, ρ is the reward/punishment sensitivity, and r_t is the outcome of that trial.

Action selection on each trial is based upon a comparison of action-weights within a squished-sigmoid, which allows the entry of irreducible noise (ξ) into action selection. This effectively prevents action selection becoming deterministic (all Go or all NoGo) even when the evidence in favour of one action over the other is comprehensively conclusive.

$$\mathrm{p}(a_t|\mathrm{s}_t) = (1-\xi)\left[\frac{\exp\left(w(a_t,s_t)\right)}{\sum_a \exp\left(w(a_t,s_t)\right)}\right] + \frac{\xi}{2}$$

Equation 3.4

Where $p(a_t|\mathrm{s}_t)$ is the probability of selection an action *a* in a state *s* at time *t*, ξ is irreducible noise, and $w(a_t, s_t)$ is the action weight for action *a* in state *s* at time *t*.

3.3.6 Model fitting

The value of six parameters (α, the learning rate; ρ_{reward}, the reward weighting; $\rho_{punishment}$, the punishment weighting; ε, the action bias; π, the Pavlovian bias, and ξ, the irreducible noise in action selection) were fitted using Expectation Maximisation (EM) algorithms, as described in [64]. The reader is referred there for full details, although we sketch out the intuition below.

EM fitting occurs iteratively. The ultimate aim is to maximise the likelihood of *all* the data given a set of population parameters that are quantified by their mean (μ) and variance (σ). This approach differs markedly from a single-subject fitting algorithm, in which data from each subject would be described by a fitted parameter, resulting in a total of 360 free parameters. This renders the procedure highly vulnerable to overfitting. Conversely, we effectively fit a total of 12 parameters, a mean and a variance for each parameter. To achieve this, on each iteration of the model, the likelihood of a value for a given subject's parameter estimate is penalised by its improbability given the current distribution of parameters in the population. This effectively 'pulls in' extreme parameter values, which might fit a single subject's data well, but are unlikely given the population distribution of parameter values. Note that we fit both groups as a single population, allowing us to compare parameters between groups using conventional statistics. Fitting each separately would be a mistake, breaking the independence of data-points in each population and precluding traditional comparisons. All parameters were fit in an unbounded space (-∞ →+∞), and then transformed to a space of either 0→+∞ (exponential transform, used for ρ, ε, and π), or 0→1 (sigmoid transform, used for α and ξ).

3.3.7 Statistical analysis

Statistical tests are described at the point of use. We used parametric tests throughout.

3.4 Results

3.4.1 Stress induction

Contrasting the stressful CPT condition with the Control condition revealed a profound effect of our manipulation upon subjective stress, autonomic, and glucocorticoid activity (Figure 3.2). Subjective stress ratings increased in the stressed group following immersion, and remained elevated above baseline throughout the task (Figure 3.2A). The manipulation did not elicit a change in subjective stress in the Control group, yielding an interaction between Group and

Time (Repeated measures ANOVA, Effect of Time: $F_{5,58}$=7.31, η^2=0.037, p<0.001; Effect of Group: $F_{1,58}$=2.01, η^2=0.021, p=0.16; Interaction: $F_{5.58}$=10.05, η^2=0.051, p<0.001). In both groups, stress at the end of the experiment was greater than at baseline (paired t-test, Stressed T_{29}=3.89, d'=0.71, p<0.001; Control T_{29}=2.64, d'=0.48, p=0.013).

Similarly, cortisol concentrations diverged between the two groups over time (Repeated measures ANOVA, Effect of Time: $F_{4,58}$=4.97, η^2=0.022, p<0.001; Effect of Group: $F_{1,58}$=2.44, η^2=0.028, p=0.12; Interaction $F_{4.58}$=9.2, η^2=0.040, p<0.001) in the manner predicted from previous work [5,7] (Figure 2B). Importantly, cortisol concentrations in the stressed group were still higher than those in the Control group at the end of the Go/NoGo task, consistent with a persistence of stress throughout the task (two-sample t-test at T30 following task completion, T_{58}=2.76, d'=0.72, p=0.0067).

Pupil diameter reflected stress induction (Figure 3.2C). Pupil diameter immediately preceding immersion was similar between groups with no difference at baseline (two-sample t-test, T_{57}=0.87, p=0.38). In the Control group, pupil diameter rapidly fell over the duration of immersion. However, the Stressed group showed a sustained elevation of pupil diameter, an effect also present in the post immersion period, producing a significant interaction between group and time (Repeated measures ANOVA, Effect of Time: $F_{2,57}$=55.61, η^2=0.017, p<0.001; Effect of Group: $F_{1,57}$=0.15, η^2=0.0025, p=0.70; Interaction: $F_{2,57}$=32.78, η^2=0.10, p<0.001). Treatment-induced changes in autonomic nervous activity were also evident in our measurements of blood pressure before and after stress induction (Table 1). Systolic and diastolic blood pressure after treatment were both higher in the Stressed group relative to Control (Systolic: Repeated measures ANOVA, Effect of Time: $F_{1,56}$=2.57, η^2=0.005, p=0.11; Effect of Group: $F_{1,56}$=1.30, η^2=0.020, p=0.26; Interaction: $F_{1,56}$=6.6, η^2=0.013, p=0.013; Diastolic: Repeated measures ANOVA, Effect of Time: $F_{1,56}$=0.13, η^2=0.0004, p=0.72; Effect of Group: $F_{1,56}$=0.08, η^2=0.0011, p=0.78; Interaction $F_{1,56}$=6.28, η^2=0.022, p=0.015). Heart rate showed a trend towards a difference between groups and an interaction with time (Repeated measures ANOVA, Effect of Time: $F_{1,56}$=9.46, η^2=0.0082, p=0.0032; Effect of Group: $F_{1,56}$=3.60, η^2=0.57 p=0.063; Interaction $F_{1,56}$=3.45, η^2=0.003, p=0.069).

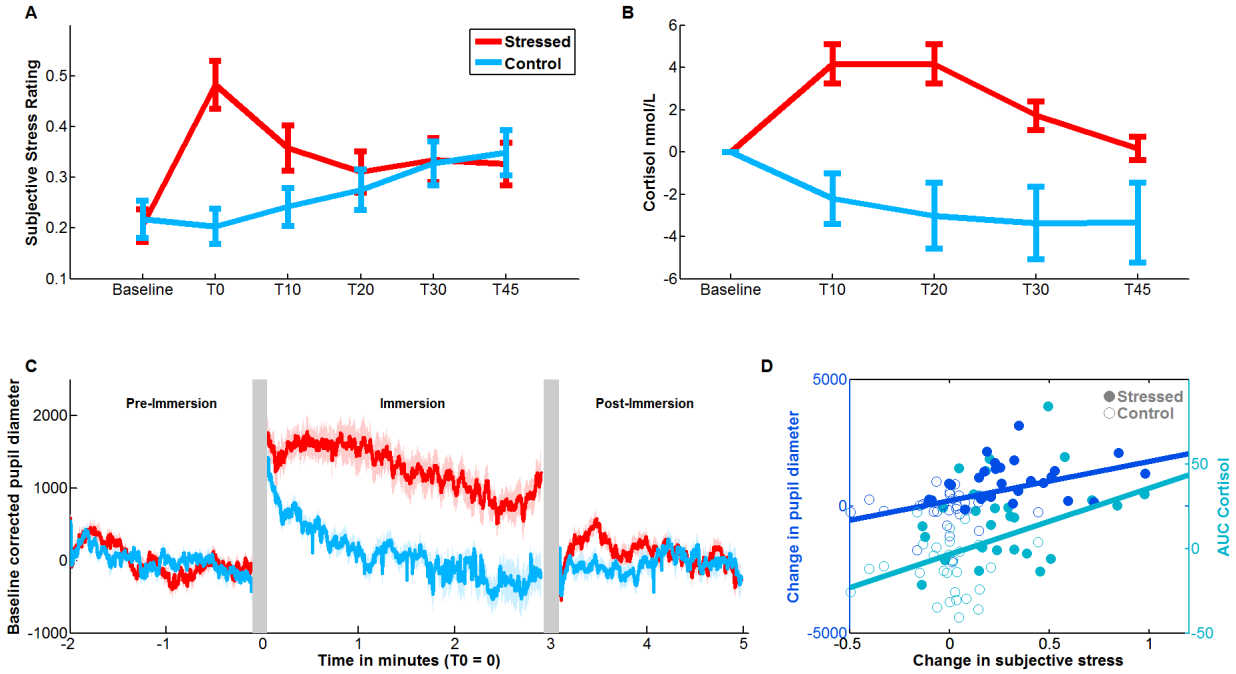


Figure 3.2 | Confirmation of stress induction (A) Subjective stress was assessed using a visual analogue scale. Immediately before timepoint T0, subjects immersed their hands in either very cold (0-1^{0}C, Stressed group) or room temperature water (25-28 ^{0}C, Control group). We observed an interaction between group and time, with an increase in subjective stress induced by the Cold Pressor Test relative to the control condition. **(B)** Salivary cortisol samples were taken at 10-15 minute intervals following immersion. We observed a robust and sustained increase in baseline-corrected salivary cortisol in the Stressed relative to the Control group, which persisted until the end of the experiment. Data baseline corrected for display. **(C)** We measured pupil diameter in a baseline period prior to T0, during immersion, and post immersion (see Figure 3.1A). Grey rectangles: signal loss during insertion and withdrawal of the hand from water. Data baseline corrected for display. **(D)** Increase in subjective stress during immersion correlated both with the increase in pupil diameter during immersion and with the Area Under Curve (AUC) of cortisol increase (see Methods). Filled circles: stressed subjects; Open circles: control subjects, with dark blue corresponding to pupil measurements and cyan to cortisol measurements. Each participant contributes two data points (one dark blue, one cyan). All error bars are SEM across participants.

Finally, we examined the relationship between our three primary stress measures (Figure 3.2D). Partial correlations indicated that manipulation-elicited change subjective stress was related to both AUC cortisol ($r_{Subjective\text{-}Cortisol}$=0.39, $p_{Subjective\text{-}Cortisol}$=0.0030) and change in pupil diameter ($r_{Subjective\text{-}Pupil}$=0.41, $p_{Subjective\text{-}Pupil}$=0.0019). However, pupillary and cortisol responses were not themselves related (r=0.056, p=0.68). This suggests that the subjective response to stressors may capture elements of the immediate catecholaminergic response reflected in pupil diameter,

whilst also predicting the extent of delayed glucocorticoid release. We note that this is a relatively rare example of concordance between multiple stress measures [24].

	Systolic Blood Pressure (mmHg)		Diastolic Blood Pressure (mmHg)		Heart Rate (Beats Per Minute)	
	Mean Before (SEM)	Mean After (SEM)	Mean Before (SEM)	Mean After (SEM)	Mean Before (SEM)	Mean After (SEM)
Stress	116.17 (2.46)	116.87 (2.42)	68.72 (1.28)	71.0 (1.41)	71.52 (1.99)	68.27 (1.79)
Control	115.48 (2.28)	110.67 (2.25)	71.38 (1.53)	69.50 (1.21)	75.41 (1.97)	74.70 (1.85)

Table 3.1 | Cardiovascular stress measures

3.4.2 Stress selectively impairs performance in learning to Go

Across all subjects, task performance was comparable to that in previous experiments [9,25]. Subjects performed best in the Go to Win condition (average correct=87.6%), with poorer performance on Go to Avoid Losing (77.0%) and NoGo to Avoid Losing (77.83%), with the worst performance observed in the NoGo to Win condition (58.3%), replicating previous studies [11]. An ANOVA confirmed the action by valence interaction predicted from extant work, along with main effects of action (participants found it easier to learn to Go than NoGo) and valence (performance was better in the Avoid Losing conditions) (Repeated Measures ANOVA, Effect of Action: $F_{1,59}=238.27$, $\eta^2=0.32$, $p<0.001$; Effect of Valence: $F_{1,59}=83.82$, $\eta^2=0.028$, $p<0.001$; Interaction $F_{1,59}=35.62$, $\eta^2=0.28$, $p<0.001$). We gathered a questionnaire measure of impulsivity (Urgency, Premeditation, Perseverance, Sensation seeking; UPPS [26]), which we hypothesized might relate to behavior in the task due to its association with dopaminergic tone [27]. This was not the case; impulsivity was not related to the number of Go responses emitted during the

experiment (r=0.19, p=0.15) or to the Pavlovian performance bias (see Methods) (r=-0.064, p=0.62).

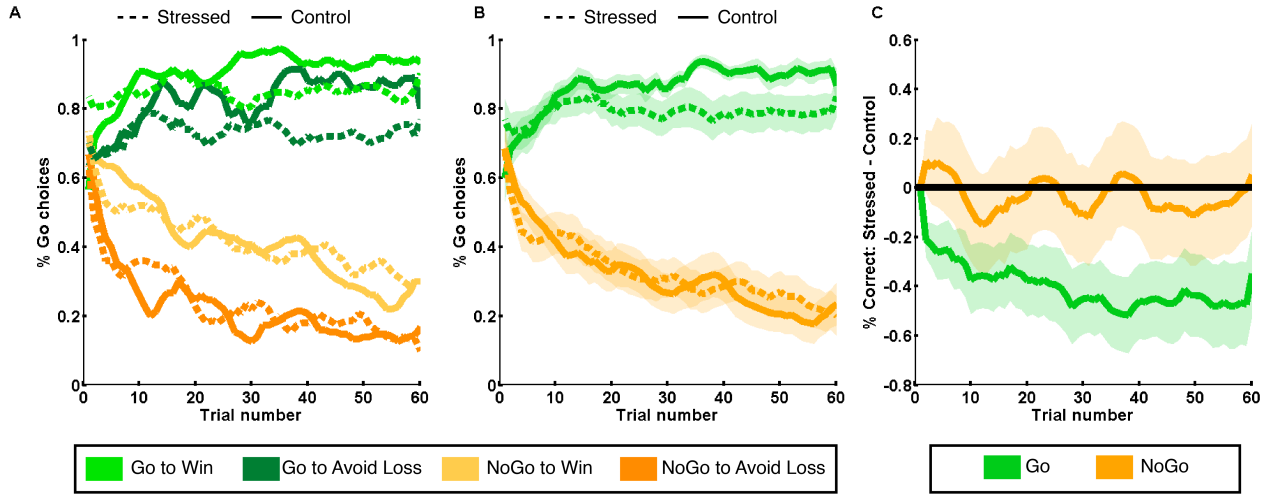


Figure 3.3 | Stress impairs learning to act (A) Percentage Go responses for each condition, for stressed subjects (dashed line) and controls (solid lines). **(B)** Grouping by Go and NoGo conditions (taking the mean across valences) reveals a deficit in Go learning in stressed subjects. **(C)** The difference in performance in Go conditions (averaged across valences) between groups, corrected for baseline differences in Go responding, grew throughout the experiment. All error bars are SEM across participants.

To visualize differences between groups and conditions, we plotted the average choices across subjects, smoothed with a 5 trial window (Figure 3.3A). Our results demonstrate a selective impairment in Go learning in the stressed group (Figure 3.3B), the magnitude of which grew over time (Figure 3.3C). Baseline-correcting for differences in Go responding at the start of the experiment in order to isolate changes in responding with learning, we tested for group differences over 10 time bins of 6 trials each. Stress decreased performance in the Go conditions at every time point (p between 0.0216 and 0.038, FDR correction for multiple comparisons applied). In order to confirm that this effect was not due to performance differences at baseline, we examined the number of correct responses made during the second-half of the experiment. The stressed group again showed reduced performance in the Go conditions (two-sample T-test, T_{58}=2.17, p=0.034). Mindful of artefacts in group-averaging [1,28], which can produce aggregate

curves suggesting incremental learning despite discrete jumps in performance at the individual level, we also performed non-continuous classification of subjects as learners and non-learners (see Methods). This allowed us to ask whether the percentage of participants that successfully learned the contingencies in each condition differed by group. This analysis also showed that stress impaired learning in Go to Win (Learners in stress group: 26, Learners in control group: 30, X^2=4.2, p=0.038) and Go to Avoid Losing (Learners in stress group: 25, Learners in control group: 30, X^2=5.45, p=0.020) but not in either of the NoGo conditions (NoGo to Win Learners in stress group: 19, Learners in control group: 18, p=0.79; NoGo to Avoid losing Learners in stress group: 26, Learners in control group: 26, p=1).

Our results suggest a specific deficit in Go-learning following stress, in accordance with the hypothesis that the aversive nature of stress leads to a bias towards inaction. Our results do not support the alternative hypothesis of a general increase in Pavlovian biases in the Go/NoGo task. Performance on the NoGo to Win condition did not differ between groups, and the decrement in the Go to Win condition is incompatible with an increase in bias, which should lead to improvement on the Pavlovian congruent conditions (Go to Win and NoGo to Avoid Losing). We further confirmed that the Pavlovian performance bias, which quantifies the difference between responding on congruent and incongruent trials (see Methods) did not differ between the two groups (, T_{58}=0.02, p=0.98, d'=0.0063).

To bolster this conclusion, we fit a reinforcement learning model that allowed us to isolate the Pavlovian interaction between action and outcome valences in trial-by-trial learning (see Methods and [2-6,25] for a description of the model). No parameters differed between groups (Learning rate T_{58}=1.54, p=0.13; Reward Sensitivity T_{58}=0.38, p=0.71; Punishment Sensitivity T_{58}=-0.99, p=0.33; Action Bias T_{58}=-.078, p=0.93; Pavlovian Bias T_{58}=1.55, p=0.12; Noise T_{58}=1.80, p=0.078) (Figure 3.4A & B). Following a recent study which found no effect of stress upon model-based learning but did observe a relationship with cortisol concentration changes [5] we looked for a correlation between model parameters and cortisol change (quantified by Area Under Curve, AUC, equivalent to the integral of cortisol changes over time) [2-7,29]. No correlations between cortisol change and model parameters were evident (Figure 3.4C) (Learning rate r=0.20, p=0.12; Reward Sensitivity r=0.097, p=0.46; Punishment Sensitivity r=-0.12, p=0.37; Action Bias r=-0.18, p=0.18; Pavlovian Bias r=0.17, p=0.20; Noise r=0.18, p=0.16). Our reinforcement learning

model did, however, provide us with trial-by-trial estimates of surprise for each subject, which we used to examine the relationship between stress and pupil responses during the task.

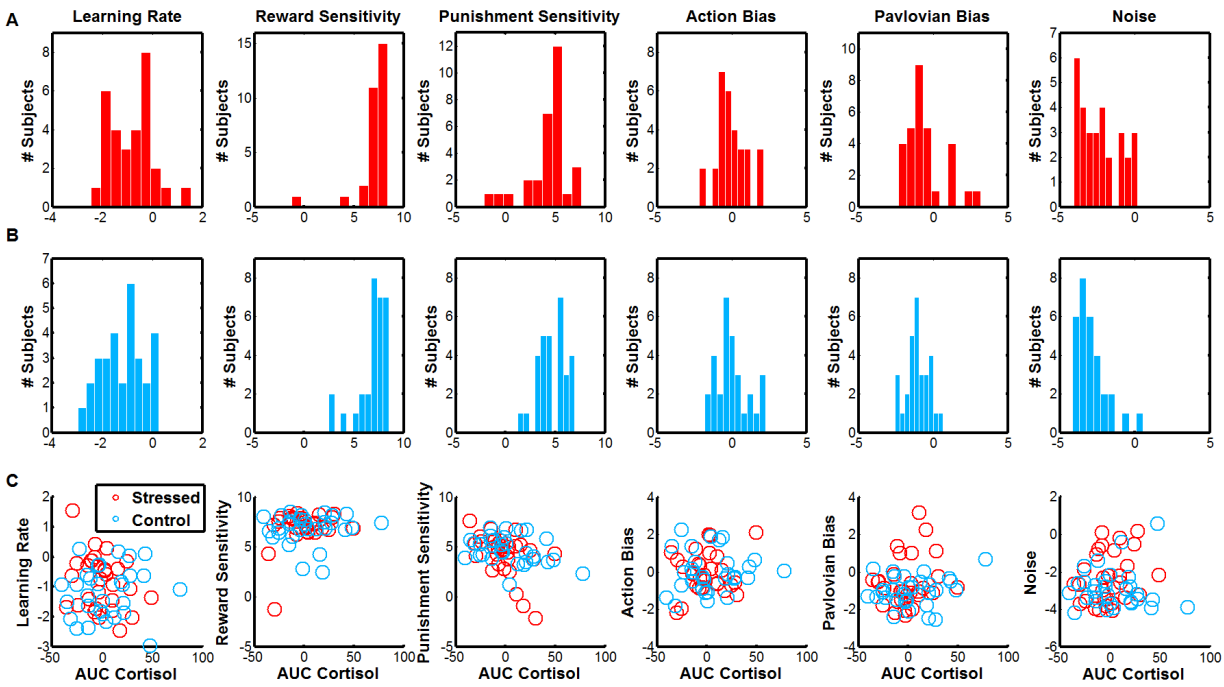


Figure 3.4 | Stress does not affect the parameters of a Pavlovian learning model (A) Distribution of parameters fit to responses from Stressed participants. **(B)** Distribution of parameters fit to responses from Control participants. **(C)** Relationship between AUC cortisol (see Methods) and model parameters. No correlations were significant. Each data point is a participant.

3.4.3 Stress alters the effect of action upon arousal

Evidence suggests that pupillary responses relate to both action and outcome processing [8-11,30-32], echoing noradrenergic responses [12,22]. Since noradrenergic dynamics are profoundly altered by stress [13,23], we asked whether stress might influence the impact of action or outcome on pupillary responses, paralleling deficits in Go learning in the stressed group.

Our task elicited reliable fluctuations in pupil diameter, which depended both upon action and valence (Figure 3.5A). Action produced a robust increase in pupil diameter, and this was further enhanced by an anticipation of loss. In light of our finding that Go learning is specifically

impaired in stress subjects, we isolated an effect of action using multiple regression models aligned to target presentation (see Methods). Since stressed subjects performed worse in the two Go conditions relative to controls, they experienced a greater number of suboptimal outcomes and, on average, larger prediction errors (due to surprising outcomes). To account for this we included outcome valence and surprise (absolute prediction error as provided by our reinforcement learning model) on each trial in the regression, along with the action taken (Go or NoGo).

Action exerted a large effect upon pupil diameter both preceding and following target presentation in both groups, as captured by large positive regression coefficients (β) (average β 1-2s following target, T_{59}=8.58, p<0.001) (Figure 5B). This captures the difference between Go and NoGo conditions depicted in Figure 3.4A. However, the magnitude of this action-induced dilatation differed between groups (Figure 3.5F). Taking the maximum β for each subject, Stressed subjects displayed a smaller increase in pupil diameter in Go vs. NoGo conditions, as captured by a smaller β relative to Control subjects (T_{58}=-2.32, p=0.024). This aligns with our behavioural observation that stress induces an action-dependent deficit in learning.

Our behavioural data suggests that stress-induced impairments are valence independent, and that Pavlovian biases within the task are unaffected by stress. To specifically examine valence and Pavlovian effects, we used an outcome-aligned regression model (see Methods), including terms for action, valence, their interaction (capturing Pavlovian biases), as well as the surprise induced by an outcome (extracted from reinforcement learning models for each participant). All three exerted systematic influences upon pupil diameter (Figure 3.5C-E). Valence (whether an outcome was positive or negative) induced both anticipatory (average β 0-1s preceding outcome, T_{59}=-3.48, p<0.001) and post-outcome (average β 1-2s after outcome, T_{59}=-7.09, p<0.001) effects (Figure 3.5C). Since valence was coded as 1, 0, or -1, corresponding to Win, No Change, or Loss respectively, the negative value of coefficients implies a larger pupil in anticipation of, and following, monetary losses compared to gains of equivalent magnitude. The size of this effect did not differ between groups (T_{58}=-0.83, p=0.41), as predicted from our behavioural finding of an action-specific, valence-independent effect of stress. The interaction of action and valence was significant post-outcome (average β 1-2s after outcome, T_{59}=-2.62, p=0.011) echoing a well-established interaction between action and valence in behaviour [9,11,14,33].

This also validates reinforcement learning models which localise this interaction to a post-outcome updating step (see Methods and [11,15]) (Figure 3.5D). In line with our behavioural and modelling results, which implied that stress did not affect the expression of Pavlovian biases within the task, there was no difference in the magnitude of this interaction between groups (T_{58}=0.25, p=0.81) (Figure 3.5F).

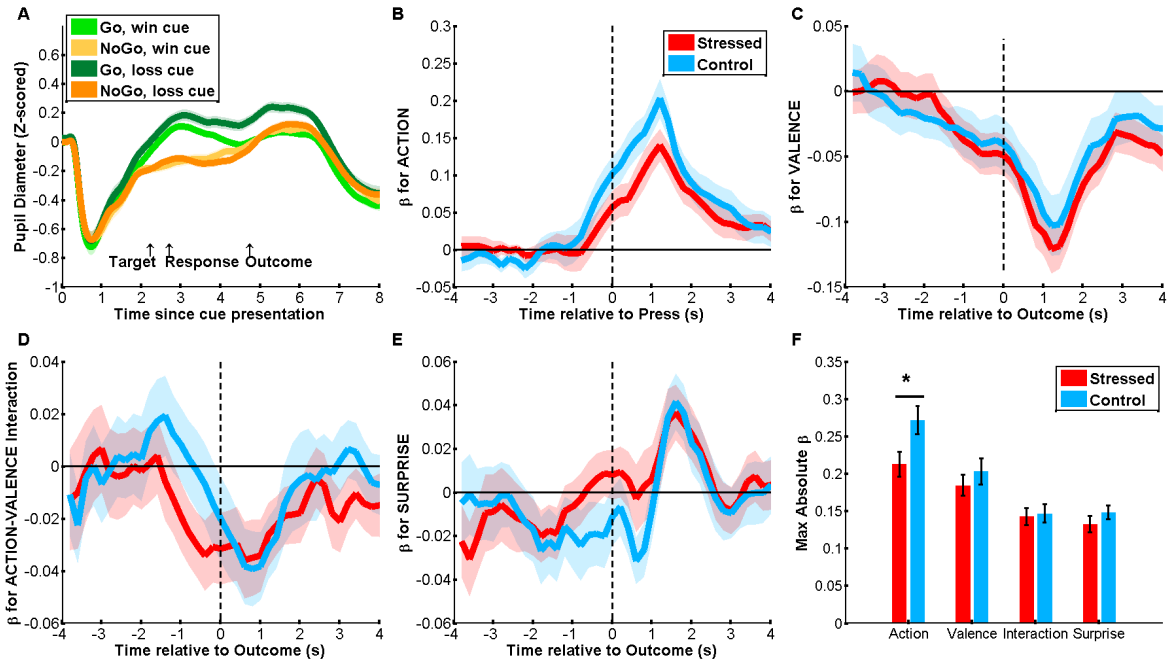


Figure 3.5 | Stress alters pupillary responses to action (A) Pupil diameter for Go and NoGo responses, separated by whether cue indicated the possibility for winning or losing money, and whether an action was produced. Action induces a robust increase in pupil diameter, amplified in trials involving the potential for monetary loss. **(B)** To isolate an effect of action upon pupil diameter, we constructed multiple regression models allowing us to account for between-group differences in outcome and surprise (see Methods). Aligned to target presentation, action exerts a large positive impact upon pupil diameter. On average, this effect was larger in the Control group (Figure 3.5F). **(C)** We used an outcome-aligned regression model (see Methods) to examine the effect of outcome valence upon pupil diameter. Valence affected pupil diameter both in anticipation of and following outcome, with negative valence increasing pupil diameter. **(D)** The same outcome-aligned regression model as in C demonstrated a significant interaction of action and valence following outcome presentation. This effect was driven by a stronger differentiation of gain and loss following Go. **(E)** Surprise increases pupil diameter. Using the same outcome-aligned regression model as in C and D, surprise (absolute prediction error) induced an increase in pupil diameter. **(F)** Action-induced pupil dilatation is reduced by stress. To avoid multiple comparisons over time, we selected the maximum absolute β for each subject for each of our 4 regressors. We then asked whether any of the effects we observed differed by group. Only the β for action was affected by stress. Coefficients for valence, action-valence

interaction and surprise did not differ between groups. All error bars are SEM across participants.

Several reports [16,30-32] have highlighted a correlation between pupil diameter and trial-by-trial estimates of surprise inferred from computational models. We quantified surprise as the absolute magnitude of the prediction errors [2,30] used to update beliefs in Q-learning models, as employed here [11,34]. We replicated previous findings that surprise exerts a positive influence upon pupil diameter post-outcome (average β 1-2s after outcome, T_{59}=2.72, p=0.0086) [19,30-32](Figure 3.5E). This effect did not differ between groups (T_{58}=-1.08, p=0.29) suggesting that the feedback signals used in error-driven learning were not altered by stress (Figure 3.5F).

3.5 Discussion

We tested two hypotheses regarding the impact of stress on learning. First, stress might induce a greater dependence upon Pavlovian biases, in line with the idea of a stress-induced general shift from computationally demanding flexible systems towards more automatic forms of control [2,5-7]. Second, an alternative account suggests that stress facilitates punishment-related behaviours, as indexed by a shift towards inaction [19-21]. We found evidence supporting this second hypothesis; stressed participants were impaired in responding to both Go cues, and showing no deficit for both NoGo cues (Figure 3.3), a conclusion supported by reinforcement learning models (Figure 3.4). This impairment in learning to act is reflected by pupillary responses in stressed subjects, who showed attenuated pupillary responses to action whilst displaying no differences in the response to outcome valence or surprise (Figure 3.5).

We note, however, that we do not observe an improvement in NoGo learning in the Stressed group, arguing against a global shift towards inaction. Stress thus appears to induce a specific deficit in action learning, whilst leaving intact the ability to learn to withhold an action. The specificity of our findings enables us to rule out several alternative explanations. Firstly, average performance on the NoGo conditions (67.6%) was lower than in the Go conditions (82.7%). Stressed subjects were therefore impaired on the *easier* of the two actions, ruling out the possibility that stressed subjects exhibit a difficulty-dependent deficit. Secondly, performance under stress in NoGo conditions was indistinguishable from the control condition, precluding a general performance deficit and underlining that stressed subjects were no more likely to correctly withhold a Go response in NoGo conditions. Stressed subjects were therefore not

merely more likely to withhold actions; they were equally likely to produce them (incorrectly) in the NoGo conditions. Thus, the deficit we describe is learning-specific and unlikely to reflect an impaired production of action by stress, but instead reflects an impairment in action learning from reinforcement. We speculatively note here that motor *excitability* is enhanced by cortisol administration [22,35], but motor *plasticity* is inhibited [23,36]. These observations underline a potentially pertinent difference between simple action production and action learning under stress.

This distinction is also important for interpreting the results of the reinforcement learning model (Figure 3.4). This model quantifies the Pavlovian bias in learning, crystallized in a parameter estimate for each participant. This parameter is affected by dopaminergic manipulations [8,9,25] and related to midline-theta activity in EEG experiments [5,7,37]. We found that this parameter was unaffected by stress, bolstering our rejection of the hypothesis that stress amplifies within-task Pavlovian biases. The model failed to capture the effect we observed, namely a selective deficit in action-learning. This is unsurprising; the model explains interindividual differences in the production of action in terms of an action bias parameter, but it cannot capture specific deficits in action *learning*. Our results suggest that expansion of the model to capture action-specific learning deficits would be fruitful.

The influence of action upon pupil diameter was attenuated in stressed subjects (Figures 3.5B & F). Pupil diameter is frequently described as an index of noradrenergic activity [20,24,30]. This link is reinforced by observations in a non-human primate study in which specific correlations were observed between pupil diameter and activity of single neurons in the noradrenergic locus coeruleus, but not the dopaminergic substantia nigra pars compacta [9,21,22,25]. Numerous reports of interactions between glucocorticoid and noradrenergic systems [11,38-41] suggest plausible substrates for altered pupillary responses under stress. Another possibility is that altered pupillary responses in the Go conditions is a *consequence* of the inability to learn. However, by including surprise as a regressor in our model of pupil diameter we show this did not account for group differences, despite an overall positive effect of surprise upon pupil diameter post-outcome (Figure 3.5E), in line with previous reports [26,30-32].

We observed an interaction between action and valence in pupillary responses (Figure 3.5D). Although fMRI studies have highlighted an interaction between action and valence in responses in the basal ganglia[9,27], we believe ours is the first physiological data with the necessary temporal precision to offer insight into the belief-updating process hypothesized to underlie this task. Computational models of learning in the Go/NoGo task place this interaction at the point at which action-weights are updated, precisely in the time window in which we observe such an interaction. Our physiological evidence for action-valence interactions following outcome thereby validates modified reinforcement learning models which incorporate Pavlovian influences in a post-outcome updating term.

We used a common laboratory stressor, the CPT, to manipulate stress levels [5-7]. This produced both an immediate increase in subjective stress and physiological arousal (Figures 3.2A & 3.2C), and a delayed, sustained increase in glucocorticoid concentrations (Figure 3.2B). Subjective stress returned swiftly to baseline following the manipulation, mirroring recent findings that subjective states such as stress reflect recent events with a relatively fast decay constant [42,43]. We do find, however, that subjective stress response to the manipulation predicts the magnitude of the sustained glucocorticoid increase that follows it (Figure 3.2D), suggesting a degree of concordance between the severity of the subjective experience of stress and its physiological sequelae, which are presumably responsible for the protracted behavioural effects we observe here.

A recent study used a reward-based paradigm in which participants performed certain actions to obtain stimulus-paired confectionary rewards from a 'vending machine' [44]. This allowed them to assess both the enhancement of a certain action given the presence of the reward-cue associated with that action (specific Pavlovian-to-Instrumental Transfer, PIT), and the general increase in responding accompanying the presence of a reward-associated cue (general PIT). They observed effects of chronic stress (assessed with a questionnaire measure) upon general transfer. Highly stressed participants did not respond more in the presence of a reward-related cue. Although the similarities between that study and this one should not be overstated – there are considerable differences between the acute stress manipulation employed here and the and assessment of chronic stress levels used there- this blunting of reward-related action production

may relate to the deficit in action-learning we observe. In both cases, stress is associated with a reduced tendency to produce actions associated with positive reinforcement.

The reported effects of stress upon learning are famously variable [45]. One potential explanation for this heterogeneity is that superficially similar behaviours are supported by distinct neural computations, which exhibit contrasting responses to stress. Recent work using choice between pairs of stimuli has suggested a shift towards the use of rewarding vs. punishing feedback during learning [46,47]. By contrast, we observe a valence-independent deficit in a task where participants choose to produce or withhold a response to a single stimulus. Such subtle differences in choice reference frame can have dramatic impact upon the neural circuits recruited during choice [48]. Specific stressors may also differ in the effects that they produce. For example, the CPT requires participants to suppress an action (the withdrawal of the hand from the ice bucket), which could conceivably prime the suppression of action-learning we observe.

Exposure to chronic uncontrollable stress induces an inability to learn to avoid future punishment, an effect described as 'learned helplessness' [49]. However, this effect also holds when the agent is required to learn NoGo responses to avoid punishment, suggesting that it is distinct from the deficit we observe here [50]. Learned helplessness is perhaps best described as a consequence of generalisation from one episode to another [51]. It is not clear how generalisation from events in the Cold Pressor Test to learning in our Go/NoGo learning task might underpin the effect we describe here, though we consider it a possibility.

What might be the functional impact of stress inhibiting learning to act? One possible explanation is that stress is typically associated with periods of high metabolic demand and uncertainty [52]. In situations where you are unsure what to do, doing nothing has the immediate advantage of being metabolically inexpensive. In the aftermath of an acute stressor such as the one we deploy here, biasing learning away from energetically costly activity may be a useful strategy for conserving resources.

In many professions, such as military, financial, or emergency medical services, sporadic surges in cortisol levels are the norm [53-55]. Our data suggest that whether an option is selected by action or inaction might have an important role in learning from decisions made under stress. One prediction is that stressed people should manifest a greater reliance upon default options, a

consequence of an impaired ability to learn from action relative to inaction. Such biases could theoretically be prevented by randomising action-outcome associations. For example, the performance of a stressed stock-trader might benefit from sometimes having selling a stock as the default option and keeping the stock requiring action.

3.6 References

1. Joëls, M. & Baram, T. Z. The neuro-symphony of stress. *Nature Reviews Neuroscience* **10,** 459–466 (2009).
2. Schwabe, L. & Wolf, O. T. Stress and multiple memory systems: from 'thinking' to "doing". *Trends Cogn. Sci. (Regul. Ed.)* **17,** 60–68 (2013).
3. Schwabe, L., Tegenthoff, M., Hoffken, O. & Wolf, O. T. Concurrent glucocorticoid and noradrenergic activity shifts instrumental behavior from goal-directed to habitual control. *J. Neurosci.* **30,** 8190–8196 (2010).
4. Schwabe, L., Tegenthoff, M., Höffken, O. & Wolf, O. T. Simultaneous glucocorticoid and noradrenergic activity disrupts the neural basis of goal-directed action in the human brain. *J. Neurosci.* **32,** 10146–10155 (2012).
5. Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences* **110,** 20941–20946 (2013).
6. Schwabe, L. & Wolf, O. T. Stress prompts habit behavior in humans. *J. Neurosci.* **29,** 7191–7198 (2009).
7. Schwabe, L., Haddad, L. & Schachinger, H. HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology* **33,** 890–895 (2008).
8. Guitart-Masip, M. *et al.* Action controls dopaminergic enhancement of reward representations. *Proceedings of the National Academy of Sciences* **109,** 7511–7516 (2012).
9. Guitart-Masip, M. *et al.* Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage* **62,** 154–166 (2012).
10. Dayan, P., Niv, Y., Seymour, B. & Daw, N. D. The misbehavior of value and the discipline of the will. *Neural Networks* **19,** 1153–1160 (2006).
11. Guitart-Masip, M., Düzel, E., Dolan, R. & Dayan, P. Action versus valence in decision making. *Trends Cogn. Sci. (Regul. Ed.)* **18,** 194–202 (2014).
12. Hershberger, W. A. An approach through the looking-glass. *Animal Learning & Behavior* **14,** 443–451 (1986).
13. Fanselow, M. S. Conditional and unconditional components of post-shock freezing. *Pav. J. Biol. Sci.* **15,** 177–182 (1980).
14. Berridge, K. C., Robinson, T. E. & Aldridge, J. W. Dissecting components of reward: 'liking', 'wanting', and learning. *Curr Opin Pharmacol* **9,** 65–73 (2009).
15. Sinha, R. Chronic Stress, Drug Use, and Vulnerability to Addiction. *Annals of the New York Academy of Sciences* **1141,** 105–130 (2008).
16. Graf, E. N. *et al.* Corticosterone Acts in the Nucleus Accumbens to Enhance Dopamine Signaling and Potentiate Reinstatement of Cocaine Seeking. *J. Neurosci.* **33,** 11800–11810 (2013).
17. Morgado, P., Silva, M., Sousa, N. & Cerqueira, J. J. Stress Transiently Affects Pavlovian-to-Instrumental Transfer. *Frontiers in Neuroscience* **6,** 93 (2012).
18. Pielock, S. M., Braun, S. & Hauber, W. The effects of acute stress on Pavlovian-instrumental transfer in rats. *Cognitive, affective & behavioral neuroscience* **13,** 174–185 (2013).
19. Robinson, O. J., Krimsky, M. & Grillon, C. The impact of induced anxiety on response inhibition.

Front Hum Neurosci **7,** (2013).
20. Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H. & Balsters, J. H. Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum Brain Mapp* **35,** 4140–4154 (2014).
21. Joshi, S., Li, Y., Kalwani, R. M. & Gold, J. I. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron* **89,** 221–234 (2016).
22. Varazzani, C., San-Galli, A., Gilardeau, S. & Bouret, S. Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. *J. Neurosci.* **35,** 7866–7877 (2015).
23. Hermans, E. J. *et al.* Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science* **334,** 1151–1153 (2011).
24. Campbell, J. & Ehlert, U. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology* **37,** 1111–1134 (2012).
25. Guitart-Masip, M. *et al.* Differential, but not opponent, effects of L -DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology (Berl.)* **231,** 955–966 (2014).
26. Whiteside, S. P. & Lynam, D. R. The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences* 669–689 (2001).
27. Norbury, A., Manohar, S., Rogers, R. D. & Husain, M. Dopamine modulates risk-taking as a function of baseline sensation-seeking trait. *J. Neurosci.* **33,** 12982–12986 (2013).
28. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences* **101,** 13124–13131 (2004).
29. Pruessner, J. C., Kirschbaum, C., Meinlschmid, G. & Hellhammer, D. H. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology* **28,** 916–931 (2003).
30. Preuschoff, K., 't Hart, B. M. & Einhäuser, W. Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience* **5,** 115 (2011).
31. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience* **15,** 1040–1046 (2012).
32. Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X. & Bishop, S. J. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience* **18,** 590–596 (2015).
33. Guitart-Masip, M. *et al.* Action controls dopaminergic enhancement of reward representations. *Proceedings of the National Academy of Sciences* **109,** 7511–7516 (2012).
34. Watkins, C. & Dayan, P. Q-learning. *Machine Learning* **8,** 279–292 (1992).
35. Milani, P. *et al.* Cortisol-induced effects on human cortical excitability. *Brain Stimulation* **3,** 131–139 (2010).
36. Sale, M. V., Ridding, M. C. & Nordstrom, M. A. Cortisol Inhibits Neuroplasticity Induction in Human Motor Cortex. *J. Neurosci.* **28,** 8285–8293 (2008).
37. Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q. & Frank, M. J. Frontal Theta Overrides Pavlovian Learning Biases. *J. Neurosci.* **33,** 8541–8548 (2013).
38. Kukolja, J. *et al.* Modeling a negative response bias in the human amygdala by noradrenergic-glucocorticoid interactions. *J. Neurosci.* **28,** 12868–12876 (2008).
39. Roozendaal, B., Okuda, S., de Quervain, D. J. F. & McGaugh, J. L. Glucocorticoids interact with emotion-induced noradrenergic activation in influencing different memory functions. *Neuroscience* **138,** 901–910 (2006).
40. Barsegyan, A., Mackenzie, S. M., Kurose, B. D., McGaugh, J. L. & Roozendaal, B. Glucocorticoids in the prefrontal cortex enhance memory consolidation and impair working memory by a common neural mechanism. *Proceedings of the National Academy of Sciences* **107,** 16655–16660 (2010).
41. McCall, J. G. *et al.* CRH Engagement of the Locus Coeruleus Noradrenergic System Mediates Stress-Induced Anxiety. *Neuron* **87,** 605–620 (2015).
42. de Berker, A. O. *et al.* Computations of uncertainty mediate acute stress responses in humans. *Nat Comms* **7,** 10996 (2016).

43. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences* **111,** 12252–12257 (2014).
44. Quail, S. L., Morris, R. W. & Balleine, B. W. Stress associated changes in Pavlovian-instrumental transfer in humans. *Q J Exp Psychol (Hove)* 1–11 (2016).
45. Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S. & Krugers, H. J. Learning under stress: how does it work? *Trends Cogn. Sci. (Regul. Ed.)* **10,** 152–158 (2006).
46. Petzold, A., Plessow, F., Goschke, T. & Kirschbaum, C. Stress reduces use of negative feedback in a feedback-based learning task. *Behav. Neurosci.* **124,** 248–255 (2010).
47. Lighthall, N. R., Gorlick, M. A., Schoeke, A., Frank, M. J. & Mather, M. Stress modulates reinforcement learning in younger and older adults. *Psychol Aging* **28,** 35–46 (2013).
48. Hunt, L. T., Woolrich, M. W., Rushworth, M. F. S. & Behrens, T. E. J. Trial-type dependent frames of reference for value comparison. *PLoS Comp Biol* **9,** e1003225 (2013).
49. Seligman, M. & Maier, S. F. Failure to escape traumatic shock. *Journal of Experimental Psychology* **74,** 1–9 (1967).
50. Minor, T. R., Jackson, R. L. & Maier, S. F. Effects of task-irrelevant cues and reinforcement delay on choice-escape learning following inescapable shock: Evidence for a deficit in selective attention. *Journal of Experimental Psychology: Animal Behavior Processes* **10,** 543–556 (1984).
51. Lieder, F., Goodman, N. D. & Huys, Q. J. Learned helplessness and generalization. *Paper presented at the Cognitive Science Conference, Berlin, Germany* (2013).
52. Koolhaas, J. M. *et al.* Stress revisited: a critical evaluation of the stress concept. *Neurosci Biobehav Rev* **35,** 1291–1301 (2011).
53. Leedy, M. G. & Wilson, M. S. Testosterone and cortisol levels in crewmen of US Air Force fighter and cargo planes. *Psychosom Med* **4** 333-8 (1985).
54. Sluiter, J. K. Medical staff in emergency situations: severity of patient status predicts stress hormone reactivity and recovery. *Occupational and Environmental Medicine* **60,** 373–375 (2003).
55. Coates, J. M. & Herbert, J. Endogenous steroids and financial risk taking on a London trading floor. *Proceedings of the National Academy of Sciences* **105,** 6167–6172 (2008).
56. Kirschbaum, C. & Hellhammer, D. H. Salivary cortisol in psychobiological research: an overview. *Neuropsychobiology* **22,** 150–169 (1989).
57. Schoofs, D., Wolf, O. T. & Smeets, T. Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. *Behav. Neurosci.* **123,** 1066–1075 (2009).
58. Bullinger, M. *et al.* Endocrine effects of the cold pressor test: relationships to subjective pain appraisal and coping. *Psychiatry research* **12,** 227–233 (1984).
59. McEwen, B. S. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews* **87,** 873–904 (2007).
60. Csikszentmihalyi, M. & Larson, R. Validity and reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease* **9** 526-36 (1987).
61. de Gee, J. W., Knapen, T. & Donner, T. H. Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences* **111,** E618–25 (2014).
62. Arakawa, H., Maeda, M. & Tsuji, A. Chemiluminescence enzyme immunoassay of cortisol using peroxidase as label. *Analytical Biochemistry* **97,** 248–254 (1979).
63. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57** 289-300 (1995).
64. Huys, Q. J. M. *et al.* Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLoS Comp Biol* **7,** e1002028 (2011).

Chapter 4: Linking computations of uncertainty to acute stress responses in humans

Previously published as:

de Berker A. O., Rutledge R. B., Mathys C., Marshall L., Cross G. F., Dolan R. J. & Bestmann S. Computations of uncertainty mediate acute stress responses in humans. *Nature Communications* **7,** 10996 (2016).

4.1 Abstract

In the **Chapter 2** we saw that fluctuations in happiness are captured by a model inspired by reinforcement learning, quantifying the expectations and rewards experienced by participants and their co-players. In this chapter we go on to capture emotional state in a dynamic environment in which participants' are continually learning about the probability of receiving a painful electric shock. Subjects learned a probabilistic mapping between visual stimuli and electric shocks. We found that Bayesian models provide a superior characterisation of learning relative to reinforcement learning models. Using the model introduced in section 1.4.5, the Hierachical Gaussian Filter, we quantified the relationship between different forms of subjective task uncertainty and acute stress responses. Subjective stress, pupil diameter, and skin conductance all tracked irreducible uncertainty. We observed a coupling between emotional and somatic state, with subjective and physiological tuning to uncertainty tightly correlated. Furthermore, the uncertainty-tuning of subjective and physiological stress predicted individual task performance, consistent with an adaptive role for stress in learning under uncertain threat.

4.2 Introduction

Stress has broad ranging physiological consequences [1]. Although acute stress is often characterized as a challenge to homeostasis, the precise features of the environment that contribute to the generation of stress responses are largely unknown. Understanding the computations that underlie acute stress responses is important for insight into how stress relates to adaptive behavior, potentially illuminating links between stress and disease and facilitating treatment of stress-related disorders [2]. In **Chapter 2** we used reinforcement learning models to capture fluctuations in happiness. Here we extend this work into the aversive domain, and show how Bayesian uncertainty estimates can be used to predict stress responses.

Extant work suggests that unpredictability and uncontrollability are central features of stressful experiences [3-7]. Classic experiments in rodents demonstrate that rats exposed to a series of electric shocks show attenuated stress responses if they are able to predict or control the timing of a stressor [8], with subsequent work in humans documenting increased pain and stress in response to unpredictable stimuli [9-11]. However, such experiments typically contrast conditions of complete unpredictability to those of complete predictability [4]. Such binary comparisons do not capture the dynamic nature of uncertainty, which varies as an organism learns about and interacts with its environment [12]. Furthermore, stress affects learning [13] suggesting that reduction of uncertainty may be a function of stress responses. Previous approaches have thus left several key questions unanswered.

Firstly, it remains to be demonstrated whether the subjective and autonomic responses to acute stressors track fluctuations in uncertainty over time, which would imply a link between the processes underpinning learning and those of stress control. Secondly, it is unclear whether such responses relate to objective unpredictability, or whether they are entrained to subjective estimates of uncertainty. If so, inter-individual variation in assessment of uncertainty might provide insight into the considerable variation in acute stress responses across individuals [14]. Finally, recent work has demonstrated that individuals track separable forms of uncertainty during learning [15], and it is unclear which form of uncertainty is important for driving subjective and autonomic responses to acute stressors [16].

To address these questions, we here adopt a subject-specific Bayesian model of decision making to track distinct forms of uncertainty and examine their relationship to acute stress responses during an aversive learning paradigm (Figure 4.1). As discussed in section 1.4.4, uncertainty can be parsed into several distinct forms [17,18], for which there exists a variety of theoretical and neurobiological evidence [12,15,19,20]. For example, forecasters predicting the performance of a football team over the coming season face several distinct sources of uncertainty when formulating their predictions. Firstly, there is 'irreducible uncertainty', which captures a chance slip by a goalkeeper or a fortuitously mis-struck shot. Irreducible uncertainty reflects the randomness inherent to any complex environment. Irreducible uncertainty might increase if it begins to rain, decreasing the accuracy with which players move. Secondly, after a series of summer signings, it may be unclear how good a team is, producing uncertainty about the probability of a team winning each match. As the season progresses, this 'estimation uncertainty' falls as the strengths and weaknesses of the team become evident. A third source of unpredictability in this context is managerial instability. Assuming the manager influences performance, uncertainty about how long the current manager will remain in charge makes it harder to predict performance. This 'volatility uncertainty' is about the stability of the context. To appreciate this, compare the stability in English football of Arsenal Football Club (1 manager for the last 18 years) to the famous volatility of Newcastle United (19 managers over the same period).

To dissect the role of these three forms of uncertainty in acute stress we utilised an hierarchical Bayesian perspective [21] (Figure 4.2A). Importantly, the model is fit individually to each subject, with two free parameters (ϑ and ω) capturing variation between individuals and allowing for divergence between subjective and objective uncertainty. In this framework, beliefs at several levels of a probabilistic hierarchy are represented as Gaussian distributions characterised by means and variances, the latter quantities corresponding to uncertainty. This naturally captures the sources of uncertainty described above: irreducible uncertainty resulting from probabilistic relationships between predictors and outcomes, estimation uncertainty resulting from imperfect knowledge of those probabilistic relationships, and volatility uncertainty reflecting potential environmental instability [17]. How these different forms of uncertainty contribute to subjective and autonomic responses to acute stressors is unknown.

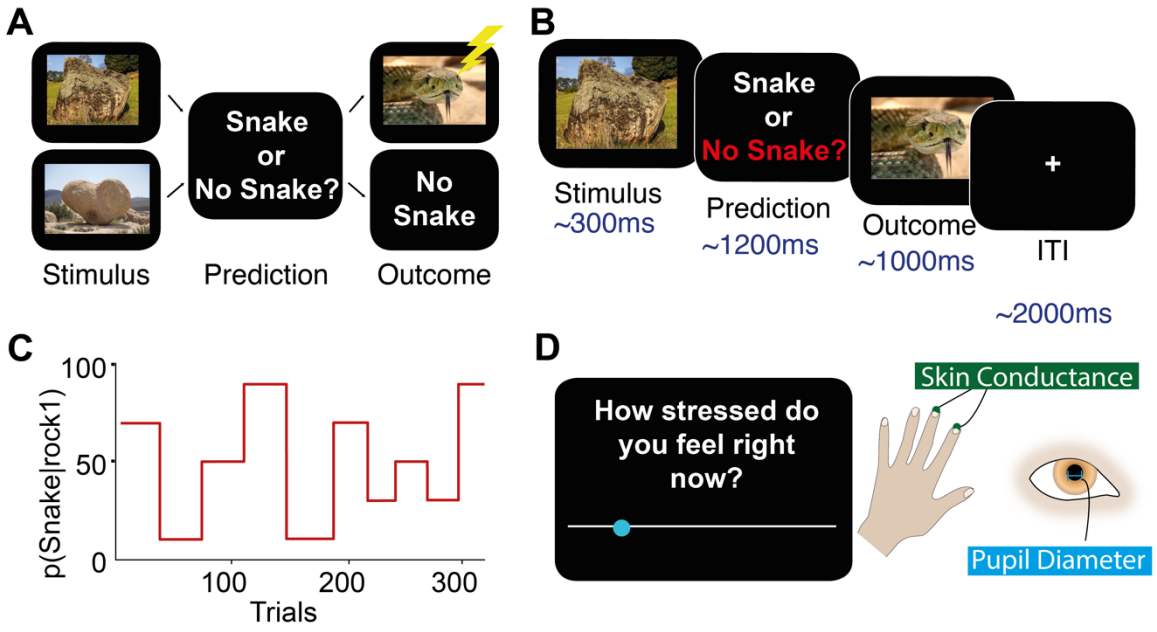


Figure 4.1 | Task structure and stress measures (A) Learning task. Visual stimuli (rocks) were probabilistically associated with outcomes (Snake or No Snake). Subjects made a prediction of the outcome on each trial. The appearance of a snake was accompanied by the delivery of a painful electric shock. **(B)** Example trial. Here the participant incorrectly predicts No Snake. Timing was jittered; see Methods. **(C)** The probabilities governing stimulus-outcome relationships shifted unpredictably over time, producing fluctuations in uncertainty. **(D)** Subjective stress ratings were collected every 4-6 trials. Measures of skin conductance (n=45) and pupil dilatation (n=22) were collected in some subjects.

We evaluated the contribution of different forms of uncertainty to subjective and physiological stress responses, using a commonly employed acute stressor over which we had precise control, electric shock [22-25]. Participants completed a probabilistic learning task (Figure 4.1) in which electric shocks were delivered with varying predictability. On each trial, a stimulus (Rock A or Rock B) was presented and participants were asked to predict whether or not there was a snake underneath (Snake or No Snake). Each time a snake was presented, participants received a painful electric shock to the hand. We used a computational model of learning [21] (Figure 4.2A) to estimate the fluctuations in uncertainty experienced by each individual, based on the predictions they made. Although the sluggish dynamics of endocrine responses preclude a detailed analysis

of the relationship between cortisol release and uncertainty [26], we measured salivary cortisol levels by way of confirmation of our stress induction. We complemented this slow measure with high-frequency assessments of subjective stress and sympathetic arousal, which vary on a time scale equivalent to the forms of uncertainty in which we are interested. This allowed us to examine the online evolution of stress responses rather than merely their delayed endocrine consequences, which may have distinct determinants and function [27,28]. As acute stress involves the coordinated action of emotional, physiological, and motivational systems [29], we measured subjective stress ratings, pupil diameter, and skin conductance (Figure 4.1D) throughout the task. Pupil diameter and skin conductance provided established measures of activity in the autonomic nervous system, a key effector of acute stress responses [30-32]. We found that all three were predicted by subjective irreducible uncertainty. We further examined inter-individual variance in the degree of coupling between uncertainty and stress responses, which we related to the ability of participants to learn in an uncertain dynamic environment.

4.3 Methods

4.3.1 Participants

All experiments were approved by the University College London Ethics Review Board. Participants (*n*=45, 25 females, aged 19-35) were recruited via the UCL Institute of Cognitive Neuroscience recruitment mailing list, and gave their written informed consent before beginning any experiments. All participants were healthy, with no history of neurological or psychiatric disorders, and no family history of epilepsy. Sample size was based on recent experiments involving stress[33].

4.3.2 Task

Participants initially underwent a shock thresholding procedure (see below). Participants then received thorough written instruction (see below) that made explicit the structure of the task, and completed a practice session of 30 trials of the probabilistic learning task used in the main experiment. They were also familiarised with the use of the subjective stress rating scale.

Our learning task was closely modelled on that used in a recent study leveraging the same computational framework in a non-stressful context [15]. Timings on each trial were jittered using

a uniform distribution to allow us to maximally divorce physiological responses to different events.

Each participant completed a set of 320 trials. On each trial, participants were presented with one of two stimuli (in our case, rocks). These stimuli remained on screen for 300ms (+/-50ms) before participants were asked to make a prediction, signalled with a button press, as to which outcome (Snake or No Snake) was likely to follow (Figure 4.1A & B). This decision was made under time pressure, with a timeout period averaging 1000ms (+/- 200ms).

Once the decision had been made, the prediction was displayed for an average of 1200ms (+/- 200ms), before the outcome was revealed. Outcomes remained on screen for 1000ms (+/- 200ms). In the case of the Snake stimulus, outcome presentation was conincident with the delivery of a shock. This was followed by an inter-trial interval of 2000ms (+/- 500ms), during which a fixation cross was displayed.

The probabilistic mapping from stimulus to outcome shifted over the course of the experiment (Figure 4.1C), requiring participants to constantly track the relationship over time. This resulted in fluctuations in the level of uncertainty. Each session of 320 trials was divided into 10 blocks of different stimulus-outcome probabilities, of lengths that varied between 26 and 38 trials. The transitions between these blocks were not made explicit to the subject. The probabilities governing each block varied from heavily biased (90/10), through moderately biased (70/30) to unbiased (50/50), allowing us to examine the effect of predictability upon stress responses. We used 4 repeats each of the biased probability blocks (2 for each bias direction i.e. 70/30 and 30/70) and 2 repeats of the 50/50 to generate 10 blocks in total.

Participants were paid a base rate of £10 and informed that they would receive an extra £0.05 for each correct prediction they made. Outcomes (correct/incorrect) were not explicitly signalled. Participants were allowed to take a self-timed break every 10 minutes.

For 41 of the 45 subjects, we report questionnaire measures of life stress (Perceived Stress Scale, PSS). Data from the remaining participants was lost due to a technical error. Some subjects (*n*=23) completed questionnaires on a separate day, whilst the remainder (*n*=22) did so after the main task. We also collected a questionnaire measure of Intolerance of Uncertainty (IUS, *n*=43) (see Supplementary Fig 4.6), depression (Beck Depression Index [BDI]), and a

questionnaire related to anxiety (Mood & Anxiety Symptom Questionnaire [MASQ])). We do not report data from the latter two questionnaires here.

4.3.3 Participant instruction

Participants were given detailed computerised instruction on the structure of the task. This emphasised that the accuracy of their predictions did not affect the number of shocks they received but did influence their earnings on the task. Understanding was confirmed with the question: "How much do you earn per correct prediction?".

We also attempted to ensure good understanding of the probabilistic relationships governing stimulus:outcome relationships. We emphasized that the probabilities were reciprocal (p(Snake|Rock A) = 1-p(Snake|Rock B)), and checked for comprehension with the question: "If the probability of a snake being under Rock A is 40%, what is the probability of it being under Rock B?"

Participants were further informed that the probabilities changed throughout the task, and that at times might appear to be random i.e. the probability of an outcome following each stimulus might be equal (50/50).

4.3.4 Shock thresholding procedure

Electric shocks were controlled using a Digitimer DS5 system in conjunction with a National Instruments Data Acquisition Board, which allowed control of shock amplitude via Matlab (Mathworks). Electrodes were placed 0.5cm apart on the first dorsal interosseous of the left hand. Electrode sites were cleaned with alcohol and a mild abrasive (NuPrep Skin Prep Gel). Shocks were delivered using BioPac Ag/AgCl electrodes filled with Sigma Spectra 360 Electrode Gel, attached using double-sided adhesive pads.

Participants first underwent a thresholding procedure which allowed us to map their subjective sensitivity to shock. Thresholding consisted of a sequence of 80 shocks, with currents of magnitudes between 0.1 and 10 mA chosen according to a staircasing procedure. After each shock, participants were asked to report how painful it was, from a rating of 1 (not painful) to 5 (very painful). We used an automated thresholding procedure inspired by Gracely et al. [62], in which separate staircases are used to estimate the transition points between each rating (1/2,2/3,3/4,4/5). For robustness, we used two independent staircases, running the QUEST

thresholding algorithm [63] for each transition. This gave a total of 8 independent staircases, with trials from each staircase randomly interleaved. At the end of thresholding, we averaged the two estimates for the 4/5 boundary to set the shock intensity for the rest of the experiment. Participants were given a sample shock at this intensity and in 3 cases we reduced the amplitude of the experimental shock by 20% at the participant's request. Shock sensitivity was also measured at the end of the task, and on average showed a slight decrease (single-sample t-test, t_{44}=2.70, p=0.097, d=0.403), equivalent to an 11% reduction in subjective pain.

4.3.5 Stress Measures.

Each participant was asked to make 65 subjective stress ratings at semiregular points throughout the probabilistic learning task (every 4-6 trials). These required participants to move a marker along a line to answer the question 'How stressed do you feel at this moment?' (Figure 1D). All analyses were conducted upon z-scored ratings to obviate between-subject differences in use of the scale.

Pupil diameter was recorded in a subset of participants (n=22) using an EyeLink 1000 System (SR Research), sampled at 100 Hz. Participants were seated in a darkened room, and asked to maintain fixation wherever possible. Stimuli were luminance matched, with the No Snake outcome signalled by a scrambled version of the Snake picture.

Skin conductance was recorded from the index and middle fingertips of the left hand using 8mm BioPac AgCl electrodes. Electrodes were filled with a 0.5%-NaCl paste (BioPac Gel 101) and attached using double-sided adhesive pads supplemented by tape. We utilised a custom recording system based upon the provision of a constant current between the two electrodes and the measurement of the resultant voltage, allowing calculation of the conductance of the skin (AT64, Autogenic Systems). This signal was converted to an optical pulse and then digitally recorded at 100Hz in Spike2 (v6.17).

In a subset of subjects (20) we collected saliva samples at 8 time points, from which we measured cortisol concentrations. In order to avoid baseline-elevation due to anticipatory stress we collected 2 baseline readings (Samples 1 & 2) on a separate day, on which participants were aware that no shocks would be received. On the day of the experiment, we collected the following samples: (3) Upon arrival (4) Immediately before task (~15 minutes after shock

thresholding) (5) 10 minutes into task (6) 20 minutes into task (7) 30 minutes into task (8) Post task. Participants salivated through straws into 2ml polypropylene tubes. Samples were frozen on the day of collection. Analysis was performed by Viapath at King's College Hospital, using a competitive immunoassay. See section 3.3.3.5 for details of assay. Some samples were not suitable for analysis due to damage in storage (28/160). To accommodate for these missing values we used a Skilling-Mack test to assess changes in cortisol over time. One subject was excluded due to baseline concentrations > 3 standard deviations away from the mean (60.59 $nmolL^{-1}$; Population Mean = 6.92 $nmolL^{-1}$; Population Standard Deviation = 13.57 $nmolL^{-1}$). To verify that this did not affect our conclusions, we repeated our analyses without excluding this subject, and found comparable results. All data was log transformed prior to analysis to render data close to normal[33].

4.3.6 Analysis of pupil diameter

Data were exported using the EDF2ASC plugin, and imported as ASCII files into Matlab. Pupil diameter measurements were down-sampled to 100Hz, and low pass filtered (4Hz, 3rd order Butterworth) [41]. Blinks were automatically detected by the EyeLink software, and removed by linear interpolation of samples 140ms either side of the blink. Data were then z-scored and detrended. No further artefact removal was necessary.

Linear modelling involved a mixture of delta and boxcar regressors convolved with a canonical pupillary response function (see below for details of response functions) [41]. For phasic responses, we also convolved regressors with the first and second derivatives of the canonical response function, a standard method in magnetic resonance imaging [65] designed to accommodate inaccuracies in the modelling of the amplitude and timing of induced responses. These were subsequently orthogonalised to their respective regressors, to apportion shared variance to the primary regressor. We took an additional step to remove signal due to changes in luminance, which produces pupil constrictions discernible from emotional/cognitive pupillary changes by their short latency [31]. We estimated a luminance-response function for each subject on the basis of passive viewing of the images in our task (each image presented 50 times, displayed for 1000ms, with a jittered inter-trial interval of between 9000 and 1100ms). This provided a response function which could then be convolved with each presentation of an image, allowing us to discriminate fast, luminance-dependent constrictions from slower

dilatations relating to cognitive variables (Supplementary Figure 4.3). Following convolution of predictors with their response functions, predictors were z-scored. Details of the full linear model can be found in Supplementary Table 4.1.

4.3.7 Analysis of skin conductance

Data were first visually inspected and 8 participants were rejected due to low recording quality.

Data from the remaining 37 participants were imported and preprocessed using tools from the SCRalyze suite [42]. Data were downsampled to 10Hz and low pass filtered (5Hz, 1st order Butterworth). We used a custom artefact rejection regime based upon the second differential of the signal; non-physiological signals were identified by their very rapid rate of change. Subjects took self-paced breaks every 10 minutes, which caused substantial changes in skin conductance amplitude due to movement of the arm, and so we discarded the first 5 trials following a break. The timeseries was then concatenated, detrended, and z-scored.

For linear models, we used a mixture of delta and boxcar regressors to represent phasic and sustained influences upon skin conductance. These regressors were then convolved with the canonical skin conductance response function outlined in ref. 35 and provided in SCRalyze (http://scralyze.sourceforge.net/) (see below for details of response functions). We utilised a variety of nuisance regressors to isolate changes in skin conductance relating to our cognitive variables of interest. All predictors were z-scored prior to modelling. Details of the full linear model can be found in Supplementary Table 4.2.

4.3.8 Modelling of learning

We modelled learning in our task using several models. Three of these (Rescorla-Wagner, Sutton K1, Hierarchical Gaussian Filter) were implemented using the HGF toolbox (http://www.translationalneuromodeling.org/hgf-toolbox-v3-0/). For a full list of priors, see Supplementary Table 4.3. For details of each model, see Supplementary Table 4.4.

The model that best explained our data was the Hierarchical Gaussian Filter (HGF) (Figure 4.2A). Introduced by Mathys et al. [21], the HGF is a Bayesian learning model not constrained by the optimality typically assumed by such models; instead, subject-specific fitting allows for inter-individual variability in learning. A recent fMRI study using the HGF highlighted its utility in assessing learning under uncertainty and its neural correlates [15]; our task and analysis was

inspired by the ones used there. For a full description of the structure of the HGF, the reader is referred to the start of the Results section and to the introduction (section 1.4.5).

The HGF was fit to each individuals' choices using Variational Bayes, with two free parameters: ϑ, a metavolatility parameter which determines step size at the third level of the HGF, and ω, which is a constant component of the learning rate at the second level.

The four quantities utilised in our analyses of stress measures are all trajectories over time, with a value that evolves according to the predictions made and outcomes experienced by that subject.

The first of these is surprise ($|\delta_1|$ in the HGF). This is the difference between the observed outcome (1 = Snake /0 = No Snake) on trial *k* and the subject's belief about the probability of that outcome:

$$|\delta^{(k)}_1| = |\mu^{(k)}_1 - s(\hat{\mu}^{(k)}_2)|$$

Equation 4.1

Where $\mu^{(k)}_1$ is the actual outcome (1 or 0) and $s(\hat{\mu}^{(k)}_2)$ is the sigmoid transformation of belief about probabilities before seeing the outcome i.e. the subject's expectations. By taking the absolute value of $\delta^{(k)}_1$ we therefore obtain a quantity which represents surprise about outcomes.

The three forms of uncertainty we consider are:

$\hat{\sigma}^{(k)}_1$: Uncertainty of predictions at the first level on trial k. Because beliefs at the first level take the form of a Bernoulli distribution, the variance is a function of the mean $\hat{\mu}_1$, namely $\hat{\mu}^{(k)}_1 * (1 - \hat{\mu}^{(k)}_1)$. This means that uncertainty has an inverted-U relationship with belief, as depicted in Figures 3C and 4B. Intuitively, this form of uncertainty represents an individual's estimate of the entropy of the environment at that moment in time; i.e. how surprising they expect things to be. We refer to it as *irreducible uncertainty*.

$\hat{\sigma}^{(k)}_2$: This is a form of informational uncertainty on trial k, representing lack of knowledge about the current stimulus:outcome relationship. Over time and in a stable environment, this uncertainty would fall to zero as the probabilities underlying the task are learned. In volatile environments, however, this is not the case. In the HGF, this form of uncertainty is approximately equivalent to a time-varying learning rate, used to update beliefs quickly when they are uncertain and slowly when they are supported by plentiful evidence. We refer to it as *estimation uncertainty*.

$\hat{\sigma}^{(k)}_3$: This can also be considered a form of estimation uncertainty, this time over the volatility of the environment at trial k. Again, it controls the speed of learning about volatility, weighting prediction errors from the probability space at the second level. We refer to it as *volatility uncertainty*.

We ran an additional pair of Rescorla-Wagner (RW) learning models to test the validity of two assumptions made by the models in the original comparison. The first is that we assume participants update probabilities symmetrically on each trial: if p(outcome|stimulus1) increases then p(outcome|stimulus2) decreases by the same amount, as constrained by our task and explained to participants. To accommodate for departures from this scheme, we tested a RW model in which the probabilities for each stimulus were updated independently. Secondly, we examined the possibility that beliefs were updated differently following trials on which shocks were delivered by fitting two learning rates (α_{shock} and $\alpha_{noshock}$) for each subject. We then compared these two models to the original RW model (in which probabilities were updated symmetrically with a single learning rate). Bayesian Model Comparison showed that the simple model comprehensively outperformed the two variants (Exceedance Probability=1). We concluded, therefore, that the assumptions of symmetric probability updating and balanced learning across Shock/NoShock trials were justified.

4.3.9 Modelling of stress

We used multiple regression models to examine the relationship between task variables, including the trajectories from the HGF outlined above, and stress responses. We used least-squared error to fit data from subjective ratings, using Matlab function glmfit. For physiological data we used robust fitting to avoid spurious fits due to unidentified artefacts. These were

implemented in Matlab with the function robustfit. In both cases we used two-tailed t-tests to assess whether parameters were different from zero i.e. whether, at the population level, a given parameter meaningfully and consistently predicted the dependent variable in question.

We compared several multiple regression models to examine the effect of uncertainty upon subjective stress. All models included parameters for the previous rating and the number of shocks received since the last rating. The third term varied between models, and captured the effects of absolute prediction error (Model 1), and uncertainty at each level (Models 2-4).

Model 1: Surprise

Model 1 omitted an explicit representation of uncertainty, but summed the absolute prediction errors ($|\delta_1|$, bounded 0-1 on each trial) since the last rating, reflecting surprise in response to outcomes given an individual's beliefs.

Model 2: Irreducible Uncertainty

Irreducible uncertainty is the variance of the Bernoulli distribution representing subject's beliefs, captured by the HGF parameter $\hat{\sigma}_1$. It is highest when $p(outcome|stim_1)$ and $p(outcome|stim_2)$ are both equal to 0.5, i.e. the sequence of outcomes is totally unpredictable. Irreducible uncertainty is also correlated with the magnitude of surprise (surprise is on average higher in uncertain situations).

Model 3: Estimation uncertainty

Uncertainty about the probabilities currently governing the observed outcomes is known as estimation uncertainty. This is represented in the HGF by the variance of the Gaussian distribution representing beliefs at the second level, $\hat{\sigma}_2$.

Model 4: Volatility uncertainty

Finally, we tested the hypothesis that subjective stress related to uncertainty at the third level, corresponding to uncertainty about the volatility of the generative process. Again, this is explicitly represented in the HGF as the variance of the Gaussian representing beliefs at the third level, $\hat{\sigma}_3$.

For physiological stress measures, we convolved predictors with canonical response functions (see below) to account for the timecourse of physiological responses [41,42]..

4.3.10 Model comparisons

We performed two sets of model comparisons. In the first we determined the best learning model to explain predictions made by the subjects, and in the second the best model for subjective stress responses. In each case, for each model and each subject, we took the model evidence (F-values for learning models or Bayesian Information Criterion [66] for multiple regression models), and used these to assess model fit by Random-Effects Bayesian Model Selection (RE-BMC) [36], as implemented in the VBA toolbox [37]. RE-BMC allows for heterogeneity in the population; the best model for each individual is allowed to vary, producing an estimate of model frequency in the population (i.e. for how many participants that model is the best model) and an exceedance probability (the probability that the model in question is the most frequently utilised in the population).

4.3.11 Response functions used in pupillary and skin conductance measures

We used linear modelling to elucidate the impact of different events upon pupil diameter and skin conductance. This approach is well-established in the fMRI literature, where the use of General Linear Model to interpret haemodynamic responses is common.

In this approach, a response function is used to describe how the output of a system (the measured variable) relates to its inputs. This response function is then convolved with a representation of the putative inputs to the system, and the resultant time course is used as a predictor in the General Linear Model. A common further step is to create secondary and tertiary regressors which are convolved with the first and second derivative of the response function [65]. These 'dispersion regressors' allow for inaccuracy in the timing or form of modelled responses; we utilize this approach in our modelling of events (this is unnecessary when the modelled process is extended over time, and represented by a boxcar).

Our response functions were drawn directly from previous studies. For skin conductance, we used the skin conductance response functions provided in SCRalyze (http://scralyze.sourceforge.net/), which are based on a gamma function convolved with a Gaussian kernel [67].

Response functions for pupillary dilatation were first discussed by Hoeks et al. [68] and we use the response function described there, which takes the form of a gamma function:

$$\text{Diameter}(t) = \text{t}^n \, e^{-n\text{t}/t_{max}}$$

Equation 4.2

The constants *n* and t_{max} dictate the shape of the response function. As advocated in the original paper, we use values of *n*=10.1 and t_{max}=930ms; we note that a recent study using this response function [41] demonstrated that this algorithm is robust to changes in the values of the constants used (see the Supplementary Information for ref. 3).

We took an additional step in our modelling of pupillary responses. As the appearance of stimuli and outcomes involved increases in luminance, they evoked light-related pupil constrictions. Importantly, such luminance responses are faster than the dilatations evoked by cognitive or emotional factors [31]. We accounted for light-related constrictions by convolving stimulus and outcome onsets with a luminance response function, which was fitted to each subject on the basis of a separate data set in which subjects were passively exposed to each of the image used in our experiment (see above and Supplementary Figure 4.3). The best fitting parameters were found using least-squared fitting implemented by the Matlab function fmincon. Calculated in this way, average *n*=3.6 and t_{max}=839ms.

4.3.12 Statistical analysis

All data analysis was completed in Matlab (Mathworks). All statistical tests were two-sided. We used one-sample t-tests to test for the significance of parameters in multiple regression models, and 2-way repeated-measures ANOVAs to analyse the time course data for skin conductance and pupil diameter. We used Pearson correlation coefficients except in the one case in which the data were *a priori* not normal, having been fit in a logit space (ϑ, Figure 2c). In this instance we confirmed non-normality using a Kolmogorov-Smirnov test, and used Spearman's Rank to assess the correlation non-parametrically. We used a Skillings-Mack test to analyse cortisol data (accounting for non-normality and missing data).

4.3.13 Code availability

Custom Matlab code for analysis of skin conductance and pupil diameter is available on request to the corresponding author. We used the HGF toolbox

(http://www.translationalneuromodeling.org/hgf-toolbox-v3-0/) for modelling of learning, the VBA toolbox for model comparison (mbb-team.github.io/VBA-toolbox/), and the SCRalyze suite for preprocessing of skin conductance data (http://scralyze.sourceforge.net/).

4.4 Results

4.4.1 Unpredictable aversive threat induces stress

On each trial, participants (*n*=45) were shown one of two rocks and asked to predict whether or not there was a snake underneath (Figure 4.1 A&B). Participants were explicitly informed of the reciprocal probabilities linking the two stimuli:

$$\mathrm{p}(\text{outcome } 1|\text{stimulus } 1) = 1 - \mathrm{p}(\text{outcome } 1|\text{stimulus } 2)$$

Equation 4.3

The probabilistic mapping from stimulus (Rock) to outcome (Snake) shifted over the course of the experiment (Figure 4.1C), requiring participants to track this relationship over time. When an outcome was revealed, the presence of a snake was deterministically associated with an electric shock delivered to the back of the left hand. Over the course of 320 trials, the probabilistic mapping between stimuli and outcomes changed every 26-38 trials, requiring participants to maintain and update their beliefs about the probability of a snake being under either rock. Participants chose correctly on 68% of trials on average. Our use of electric shock proved an effective elicitor of cortisol release. A Skillings-Mack test (used to account for non-normality and missing data, see Methods) confirmed that cortisol concentrations changed over the course of the experiment (Skilling's Mack T_7=18.48, p=0.010). Paired tests indicated that this was due to an elevation of cortisol above baseline 20 minutes after the first shocks were received (Wilcoxon rank sum, Z=2.20, p=0.028), in line with the typical timecourse of endocrine responses [26 28,33].

4.4.2 Hierarchical Bayesian learning explains predictions of shock

We compared the performance of three learning models in explaining the predictions that participants made on each trial, defining our model space by reference to a recent study using a similar prediction paradigm [15]. The simplest was a Rescorla-Wagner model [34], in which beliefs are updated by prediction errors with a fixed learning rate. Our second model, the Sutton-K1 [35], allows the learning rate to vary as a function of recent prediction errors. The third model was a 3-level Hierarchical Gaussian Filter (HGF) [21] in which beliefs are updated via prediction errors, with learning rates influenced by uncertainty about the veracity of current beliefs and environmental stability (Figure 4.2A). The HGF is hierarchical in the sense that learning occurs simultaneously on multiple levels. We consider a three level model, as this has been shown to

describe learning in a similar task where tone-picture associations were learned in a non-stressful context [15] (see introduction section 1.4.5).

The first level of the HGF constitutes participants' predictions for each trial, the second level represents beliefs about probabilities that give rise to those predictions, and the third level quantifies the estimated volatility of the probabilities. On each trial, the model provides an estimate for each level, before the outcome is revealed and the estimate updated accordingly. Circumflexes (^) are used to distinguish the pre-update estimates from the updated version. The model is Gaussian in that predictions at each level are represented by a Gaussian distribution, described by its mean, $\hat{\mu}_i$, and variance $\hat{\sigma}_i$, with i denoting the level in question (1, 2, or 3 in our model). The variance $\hat{\sigma}_i$ represents the uncertainty of the estimate at each level. As previously alluded to, the first, second, and third level variance ($\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\sigma}_3$) correspond to irreducible, estimation, and volatility uncertainty respectively. Updates of beliefs at each level occur via prediction errors that propagate upwards and are weighted by the relative uncertainty of the level that generated them to the uncertainty of the level being updated, a form of precision-weighting [15].

We compared these three models (Rescorla-Wagner; Sutton K1; HGF) in a Random-Effects Model Comparison [36], using tools available online [37] (Figure 4.3A). We found that the HGF was the best model by a considerable margin (Model Frequency=82%, Exceedance Probability~1). This is a close replication of the aforementioned study in which tone-picture associations were learned in a non-stressful context [15]. Having ascertained that the HGF was the model that best explained the predictions made by our participants, we proceeded to examine the distribution of fitted model parameters across the population. Fitting of the HGF allows for variance between individuals [15,21], which in this instantiation is expressed by two parameters: ω and ϑ. ω is a constant component of the learning rate at the second level, capturing interindividual variability in how rapidly people update their beliefs. ϑ determines the rate of update of the third level; this parameter can be understood as capturing 'metavolatility', with higher values implying a belief in a less stable world (Figure 4.2A).

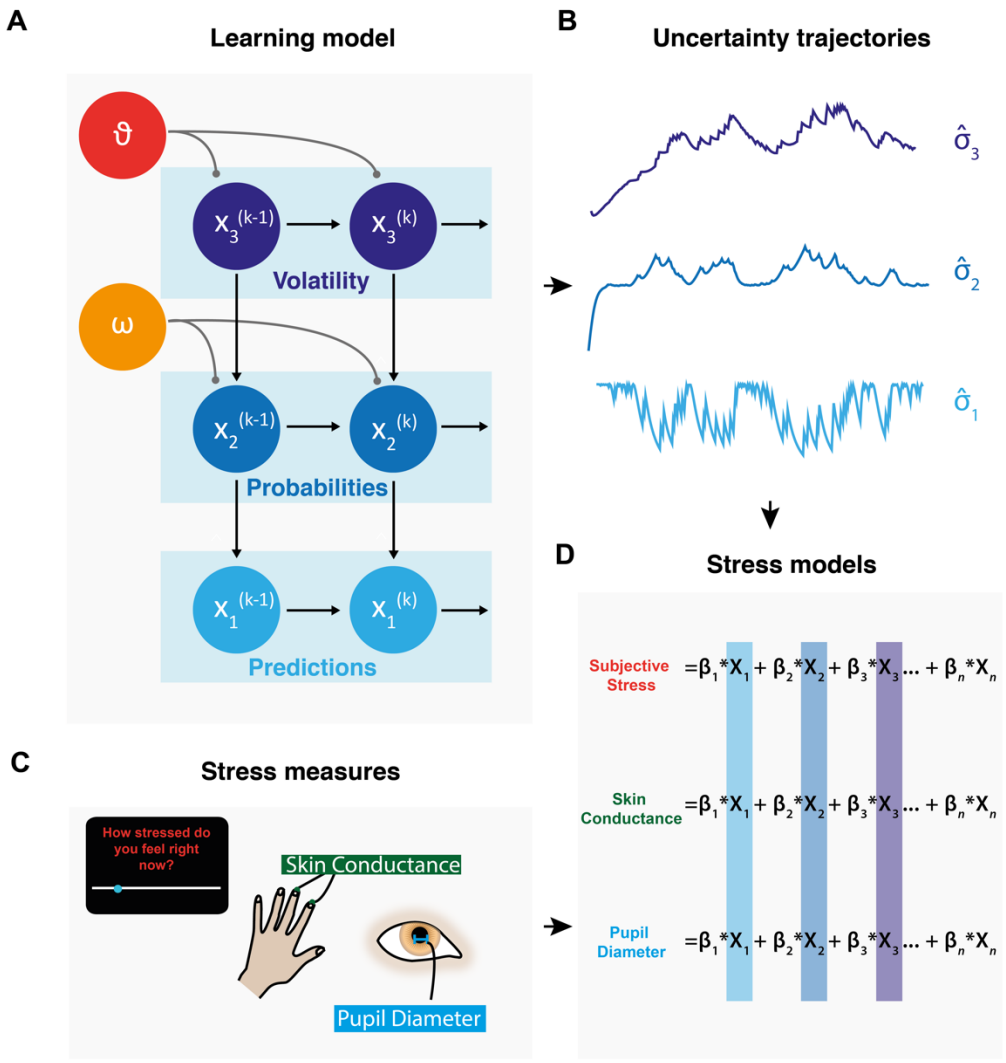


Figure 4.2 | Modelling of learning and stress (A) Hierarchical Gaussian Filter (HGF) model [21]. Beliefs are represented in probability distributions organized in a hierarchy, with the speed of updating at each level influenced by the estimate at the level above. This allows learning to occur more quickly in volatile environments. Each level is Gaussian, characterised by a mean (μ) and a variance (σ), which corresponds to uncertainty. These representations unfold over time, with the model furnishing an estimate at each level, for each trial. We take these dynamic representations of uncertainty from this model **(B)** and use them to predict stress responses **(C)** using linear modelling **(D)**. The resultant regression coefficients ($\beta_{1\text{-}n}$) quantify the influence of each form of uncertainty upon that stress measure. Note that $X_{1\text{-}n}$ are the regressors in the linear model, which include but are not limited to uncertainty trajectories; we also include terms such as number of shocks, and nuisance variables such as gaze co-ordinates.

We found that individuals' metavolatility parameter correlated with levels of chronic stress, as assessed by a questionnaire measure of life stress, the Perceived Stress Scale [38] (Figure 4.3B) (Spearman ρ=0.39, *p*=0.014; non-parametric statistics used due to non-normality of ϑ [Kolmogorov–Smirnov test, *p*<0.001])). This suggests that people who report higher levels of life stress behave as if they believe the environment is more uncertain, indicating that chronic stress levels may be affected by prior exposure to environments of high uncertainty. This confirms that interindividual variability in stress relates to variability in the beliefs about uncertainty, as expected if stress responses are tuned by exposure to uncertainty in the real world. Having established a relationship between beliefs about uncertainty and a static subjective measure of chronic stress, we next addressed the coupling between dynamic stress measures and the trajectory of uncertainty within the task.

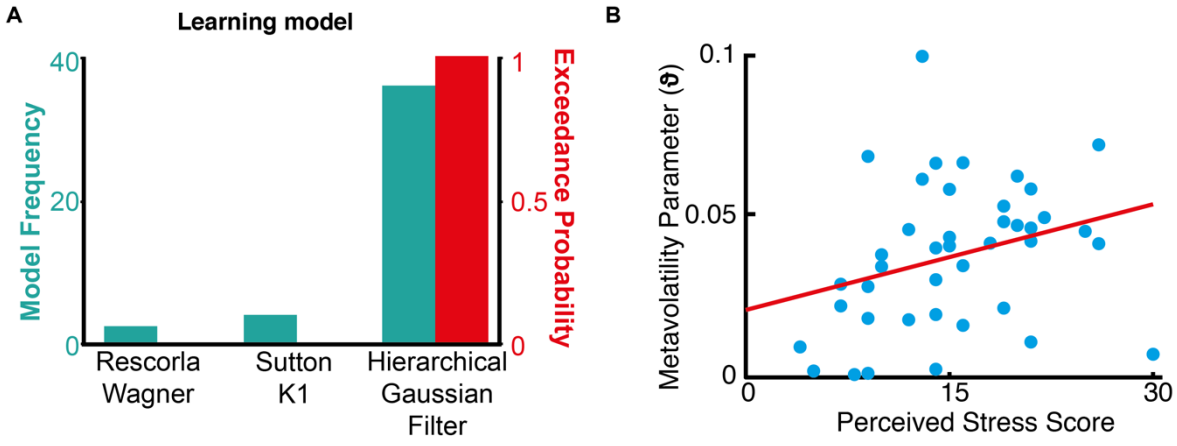


Figure 4.3 | Assessing models of learning (A) Random-effects Bayesian Model Comparison confirmed that the HGF outperformed fixed-learning-rate models (Rescorla-Wagner) and variable-learning-rate non-Bayesian models (Sutton K1). **(B)** Life stress was assessed with a Perceived Stress Scale [38]. Life stress scores were correlated with the metavolatility parameter (ϑ) in the HGF, suggesting that more stressed individuals believe the world to be less stable (n=41; Spearman ρ=0.38, p=0.014).

4.4.3 Current irreducible uncertainty predicts subjective stress

Having found that the HGF was the appropriate model of learning in our task, we asked how the dynamic quantities represented in this model related to acute stress responses. In fitting the

HGF to each participant, we obtained estimated trajectories of surprise (absolute prediction errors) and uncertainty over time (Figure 4.2B). To assess the influence of surprise and uncertainty upon subjective stress, we fit multiple regression models (Figure 4.2D) to subjective stress ratings for each participant (example trajectory shown in Figure 4.4A). We compared the ability of four different models to predict stress ratings. All four models incorporated the previous stress rating and the number of shocks received since the last rating. Our predictions therefore took the following form:

$$\text{Rating}(k) = \beta_1 \cdot \text{Rating}^{(k-1)} + \sum_{\text{i}(k-1)}^{\text{i}(k)} \beta_2 \cdot \text{Shocks}^{(k)} + \beta_3 \cdot \text{Variable Term}^{(k)}$$

Equation 4.4

Here *k* represents the rating number [1-65], and *i(k)* is the trial number [1-320] associated with rating *k* (ratings occurred on average every 4-6 trials). Hence the shock term is the number of shocks received since the last subjective stress rating.

Our first model summed the surprise experienced by a participant since the last rating, as captured by the variable δ_1 in the HGF. The other three regression models quantified the uncertainty represented at each level of the HGF: irreducible uncertainty ($\hat{\sigma}_1$), estimation uncertainty ($\hat{\sigma}_2$), and volatility uncertainty ($\hat{\sigma}_3$).

Subjective stress responses were best predicted by a model incorporating solely the current level of irreducible uncertainty ($\hat{\sigma}_1$) (Model Frequency=41%, Exceedance Probability=0.849, Figure 4B). The resultant model is depicted in Figure 4.4D. As predicted, participants reported being most stressed when they believed the current state was high in irreducible uncertainty. All parameters were significantly greater than zero (Single-sample t-tests on parameters from multiple regression: Previous Rating β=0.25, t_{44}=7.60, p<0.001; Shocks β=0.074, t_{44}=3.58, p<0.001; Irreducible Uncertainty β=0.099, t_{44}=3.22, p=0.0024). We found no evidence for a model of subjective stress featuring multiple forms of uncertainty (Supplementary Figure 4.1A). Additionally, we found that the estimates of subjective irreducible uncertainty furnished by the HGF provided better predictions of subjective stress than the objective irreducible uncertainty on each trial (Supplementary Figure 4.1B and 4.1C).

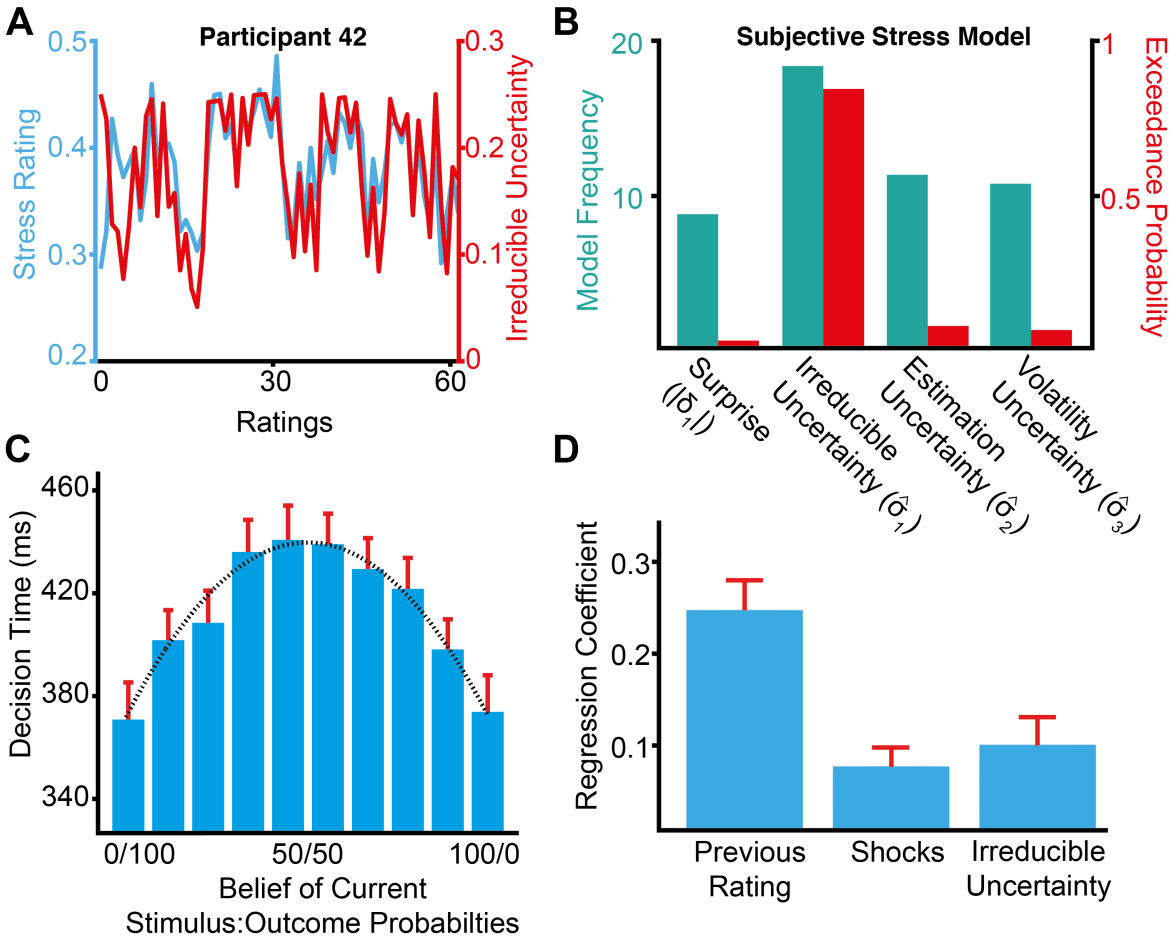


Figure 4.4 | Irreducible uncertainty predicts subjective stress (A) Example subjective stress trajectory for one participant (blue) and irreducible uncertainty estimates for that individual (red). **(B)** Regression models of subjective stress. All models shared two components: the value of the previous rating, and the number of shocks delivered since the last rating. The surprise model summed the surprise ($|\delta_1|$) for each outcome since the previous rating, whilst models $\sigma_{1\text{-}3}$ included the estimated uncertainty at each level of the HGF at the time of rating. Irreducible uncertainty (σ_1) provided the best fit to our participants (n=45). **(C)** Irreducible uncertainty predicts prediction response times. A curve describing the variance of a Bernoulli distribution representing beliefs about probabilities, corresponding to irreducible uncertainty, predicts average response times (Pearson r=0.99, p<0.001). **(D)** The winning regression model predicting subjective stress responses (mean r^2=0.25). Shocks and irreducible uncertainty both predicted subjective stress ratings (single-sample t-tests, p<0.001; p=0.0024). Error bars are SEM.

Subjective irreducible uncertainty is highest in our task when the subject's estimated probability of a shock is 50%, corresponding to a situation where the environment is utterly unpredictable, and maximal in entropy [17,18]. There is an inverted-U relationship between irreducible uncertainty and probability, according to the variance of a Bernoulli distribution (Uncertainty = Probability x (1-Probability)). This relationship was also reflected in participants' behavior, in that they were slowest making decisions under conditions of maximal uncertainty (Figure 4.4C) (Pearson correlation, r=0.99, p<0.001).

We found that subjective stress responses are predicted by the trajectory of irreducible uncertainty experienced by each individual. The link between subjective and physiological indices of stress is problematic, with proposals that stress responses should exhibit coherence [29] not well supported by extant data [39]. Consequently, we next asked whether irreducible uncertainty also predicts physiological arousal, examining its relationship with two standard physiological stress measures, pupil diameter and skin conductance (Figure 4.2C).

4.4.4 Physiological stress reflects uncertainty and surprise

As a first step in gaining insight into the role played by uncertainty, we epoched physiological responses (pupil diameter, n=22; skin conductance, n=37) by trial, starting 2 seconds before an outcome was revealed. Based on evidence that pupil diameter and skin conductance reflect surprise [40], and building upon our finding that subjective stress responses are predicted by irreducible uncertainty, we implemented median splits, separating trials according to whether they were high or low in irreducible uncertainty and high or low in surprise. This resulted in four groupings (high/high, high/low, low/high, low/low). Taking the average across participants, we observed that uncertainty increased pupil diameter throughout the trial (Figure 4.5A), with an additional, positive effect of surprise approximately 2 seconds after outcome presentation. The time-course of skin conductance responses was similar, albeit slower (Supplementary Figure 4.2A). Two-Way ANOVAs demonstrated that both pupil diameter and skin conductance were increased by irreducible uncertainty (Pupil: $F_{1,21}$=22.56, p<0.001, η^2=0.051; Skin Conductance $F_{1,36}$=9.36, p=0.004, η^2=0.104) and surprise (Pupil: $F_{1,21}$=20.71, p<0.001, η^2=0.045; Skin Conductance: $F_{1,36}$=12.40, p=0.001, η^2=0.070), with no interaction (Pupil: $F_{1,21}$=0.51, p=0.48; Skin Conductance: $F_{1,36}$=0.14, p=0.71).

Within the framework of our model, information about uncertainty on the current trial is available to the subject before the trial begins, as it is computed on the basis of trial history [21]. Consequently, we asked whether baseline pupil diameter and skin conductance, as assessed at the start of each trial, reflected the subjective belief of probabilities on that trial, as represented in our learning model. We found that baseline arousal displayed an inverted-U relationship with belief, strikingly reminiscent of the relationship between reaction time and belief (compare Figure 4.4C and 4.5B). To confirm this relationship, we show that a curve fit to the variance of a Bernoulli distribution, describing the relationship between irreducible uncertainty and belief, captured this relationship well (Pearson correlations, Pupil: r=0.96, p<0.001; Skin Conductance: r=0.84, p=0.002).

To examine more precisely this relationship between uncertainty, surprise, and skin conductance, we employed a model-based approach (Figure 4.2D), convolving response functions for pupillary [41] and skin conductance [42] responses with our predictors (see Methods for full details of model and Supplementary Figure 4.3 for details of pupillary response function). We included surprise as a regressor in our models to ensure that responses to uncertainty were independent of the surprise at outcome (Supplementary Figure 4). We found that the level of irreducible uncertainty on each trial was a significant predictor of both pupil diameter (robust regression β=0.11, single-sample t-test t_{21}=4.72, p<0.001) and skin conductance (β=0.044, t_{36}=2.25, p=0.031) (Figure 4.5C).

Finally, we asked whether the sensitivity of physiological stress to uncertainty related to that inferred from reported subjective stress responses. We took the magnitude of the regression coefficients (β) for irreducible uncertainty from our models of pupil diameter and skin conductance and compared them to the equivalent terms from our subjective stress model. In both cases, the two were positively correlated (Pearson correlation, Pupil: r=0.52, p=0.013; Skin Conductance: r=0.38, p=0.021) (Figure 4.5D) such that individuals whose subjective reports were more sensitive to uncertainty also showed a greater impact of uncertainty upon their physiological stress responses. This concordance between emotional and physiological state is predicted by theories of emotion [43], although direct evidence for this relationship is rare [44]. Our computational perspective on the cognitive dynamics of stress responses reveal a strong

coherence between emotional and physiological systems, though we note that the sensitivity of the two physiological measures were not themselves correlated (Supplementary Figure 4.4).

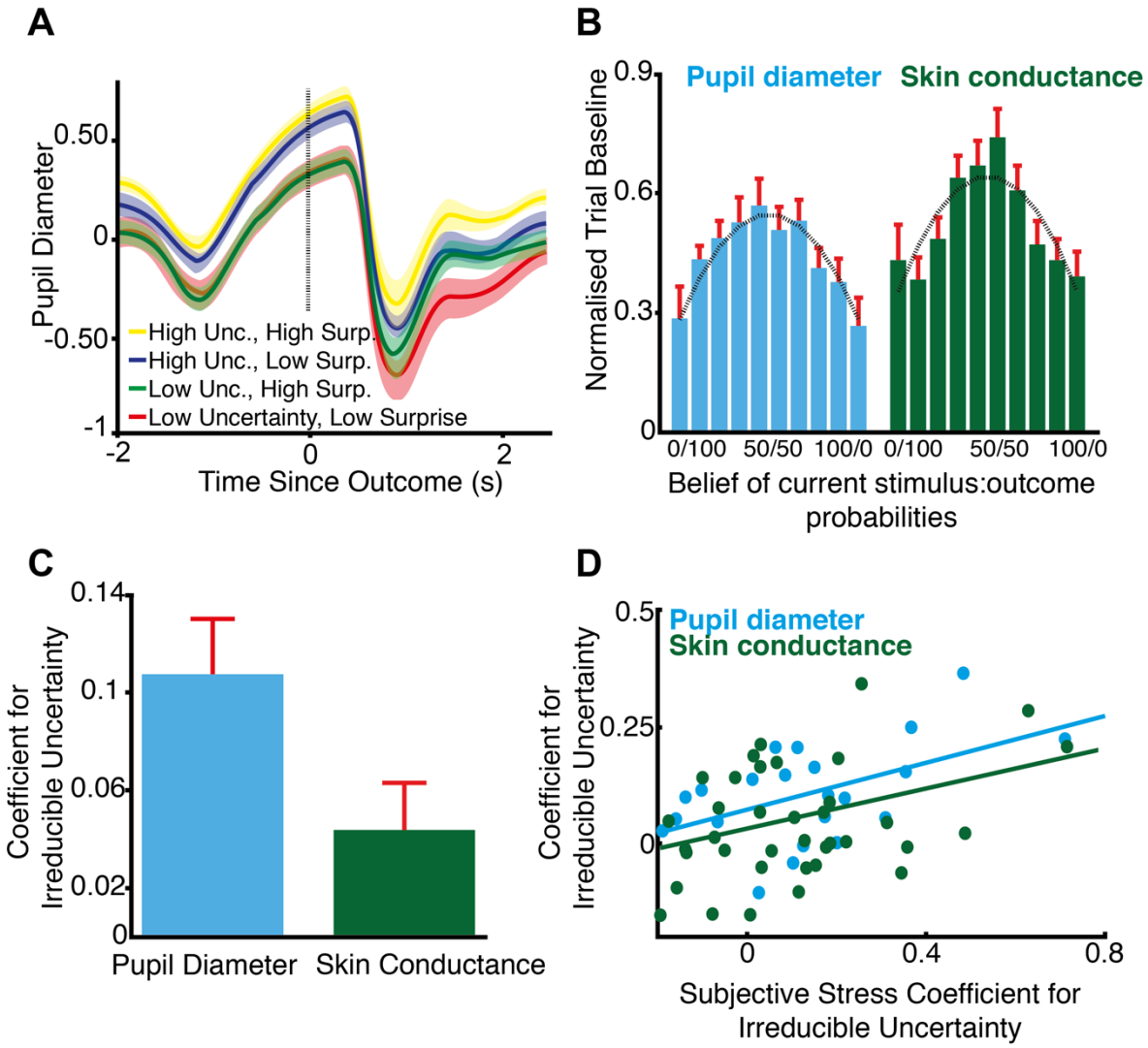


Figure 4.5 | Physiological responses reflect uncertainty and surprise (A) Median splits indicate that both irreducible uncertainty and surprise increase pupil diameter (both $p<0.001$). Shading is SEM across participants. **(B)** Baseline pupil diameter and skin conductance on each trial displayed a clear inverted-U relationship with belief, as seen for subjective stress. A curve describing the variance of a Bernoulli distribution fit well (Pearson correlations: Pupil diameter: $r=0.96$, $p<0.001$; Skin conductance: $r=0.84$, $p=0.02$). Error bars represent SEM. **(C)** A multiple regression model demonstrated a role for uncertainty throughout the trial for both pupil diameter (robust regression $\beta=0.11$, single-sample t-test, $t_{21}=4.72$, $p<0.001$) and skin conductance (robust regression $\beta=0.044$, single-sample t-test, $t_{41}=2.25$, $p=0.031$). Error bars represent SEM. **(D)** Across subjects, the sensitivity of subjective stress to irreducible uncertainty correlated with the sensitivity observed in pupil diameter (Pearson correlation, $n=22$, $r=0.53$,

p=0.014) and skin conductance (Pearson correlation, *n*=37, *r*=0.38, *p*=0.021) models. Each data point is one participant.

4.4.5 Uncertainty-tuning of stress predicts performance

We hypothesised that if the tuning of stress responses to uncertainty is adaptive, the degree of coupling between uncertainty and stress responses would predict how well participants performed in our task. We found this was indeed the case, as both subjective (Pearson correlation, *r*=0.37, *p*=0.012) and pupillary (Pearson correlation, *r*=0.62, *p*=0.0023) sensitivity to uncertainty predicted task performance (Figure 4.6A and B). Thus, the degree to which stress responses track irreducible uncertainty in the environment predicts learning under uncertain threat, in accordance with an adaptive account of stress responses under uncertainty. No such relationship was evident between gross measures of stress such as the mean or standard deviation of stress ratings, highlighting the utility of our model-based approach (Supplementary Figure 4.5). We also observed a negative relationship between intolerance of uncertainty and pupil diameter (Supplementary Figure 4.6).

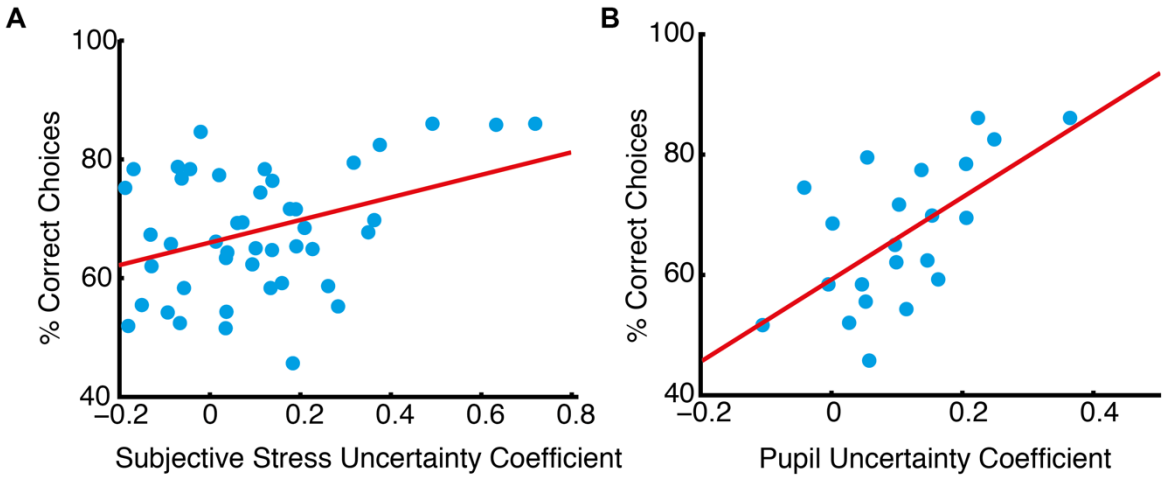


Figure 4.6 | Relationship between uncertainty sensitivity and task performance (A) Subjective stress sensitivity (the regression coefficient for uncertainty in the subjective stress model) correlated with how frequently participants predicted the correct outcome (Pearson correlation, n=45, r=0.37, p=0.012). **(B)** Pupillary sensitivity to uncertainty also predicted performance (Pearson correlation, n=22, r=0.62, p=0.0023). Each data point is one participant.

4.5 Discussion

Stress responses are co-ordinated physiological and behavioral responses to environmental challenges [1]. The precise features of the environment that generate stress have proved hard to pin down, particularly within a quantitative framework. Here we reveal a strong relationship between stress and subjective estimates of a quantifiable property of the environment, namely irreducible uncertainty. This demonstrates that computational models of learning can provide quantitative metrics of environmental and psychological variables that drive emotional and physiological stress responses. In the present case this highlights a striking relationship between a specific form of uncertainty and stress responses.

4.5.1 Understanding emotion using learning models

We built upon recent progress in computational modelling of subjective well-being [45], to inform a dissection of subjective stress responses. This was made possible by a hierarchical Bayesian model which allowed us to infer the trajectory of uncertainty experienced by each individual in our experiment. The use of computational models has proved indispensable in exploring the relationship between stress, genotype, and behaviour [46], but has not to our knowledge previously been applied to understand the genesis of stress responses. Our finding that such models can be used to link subjective uncertainty and stress responses adds to a growing consensus that detailed quantitative models are indispensable for the understanding of complex biological and mental phenomena [47-50].

4.5.2 Linking subjective and physiological stress responses

Having identified that irreducible uncertainty best predicted subjective stress responses, we next asked whether physiological responses were similarly predicted by uncertainty. Pupil diameter is a readout of central arousal thought to relate to noradrenergic activity in the locus coeruleus [51]. Recent evidence suggests that locus coeruleus firing correlates with pupil diameter in the macaque monkey, and that stimulation of the locus coeruleus is sufficient to induce changes in pupil diameter [52]. Although noradrenergic dynamics are crucial in orchestrating acute stress responses [53] as well as their behavioral [54] and mnemonic [55] impact, pupillometry is not typically employed in studies of stress (though see Henckens et al. (2012) for a notable exception). This is a potentially rich avenue of exploration, given that pupillary response metrics provide insight into emotional [31] and cognitive dynamics [41,56,57]. Our finding that pupil diameter reflects the

surprise associated with an outcome, regardless of valence, replicates previous results [56-58]. Furthermore, we also find a correlation between pupil diameter and the current level of irreducible uncertainty, with greater pupil diameter associated with higher levels of uncertainty. In rewarding environments, pupil diameter has been shown to reflect estimation uncertainty and, as we find here, irreducible uncertainty, often referred to as risk [56,57].

A recent study examining individual differences in aversive learning found a post-outcome pupillary sensitivity to volatility [58]. The authors found that learning-rate malleability in the face of changing volatility was related to trait anxiety. We additionally establish a link between chronic stress states and beliefs of environmental volatility (Figure 4.2C). However, the stimulus-outcome contingencies used in the previous study kept irreducible uncertainty roughly constant, precluding the comprehensive characterisation of multiple forms of uncertainty and the relation to the dynamics of emotional and physiological stress responses that we perform here. Our finding that pre-outcome pupil diameter correlates with irreducible uncertainty, and that this modulation is proportional to the effect of uncertainty upon subjective stress, is uniquely enabled by our design, and not inconsistent with the volatility-sensitivity observed previously.

Fluctuations in skin conductance depend upon activity in the sympathetic nervous system [42], a key component of physiological stress responses [53]. Our finding that skin conductance tracks uncertainty chimes with findings using the Iowa Gambling Task, in which participants make choices between decks of cards that produce rewards and punishments of varying magnitudes. Greater skin conductance responses are elicited whenever participants choose a card from the pack with higher risk, as defined by the variance of the outcome distribution [59]. However, binary choices in static environments do not reveal whether skin conductance truly reflects uncertainty, or instead some aspect of the decision process. Our results suggest that even in non-instrumental settings, somatic state relates closely to uncertainty in the environment. Furthermore, we show that these responses are dynamically driven by evolving internal estimates of irreducible uncertainty.

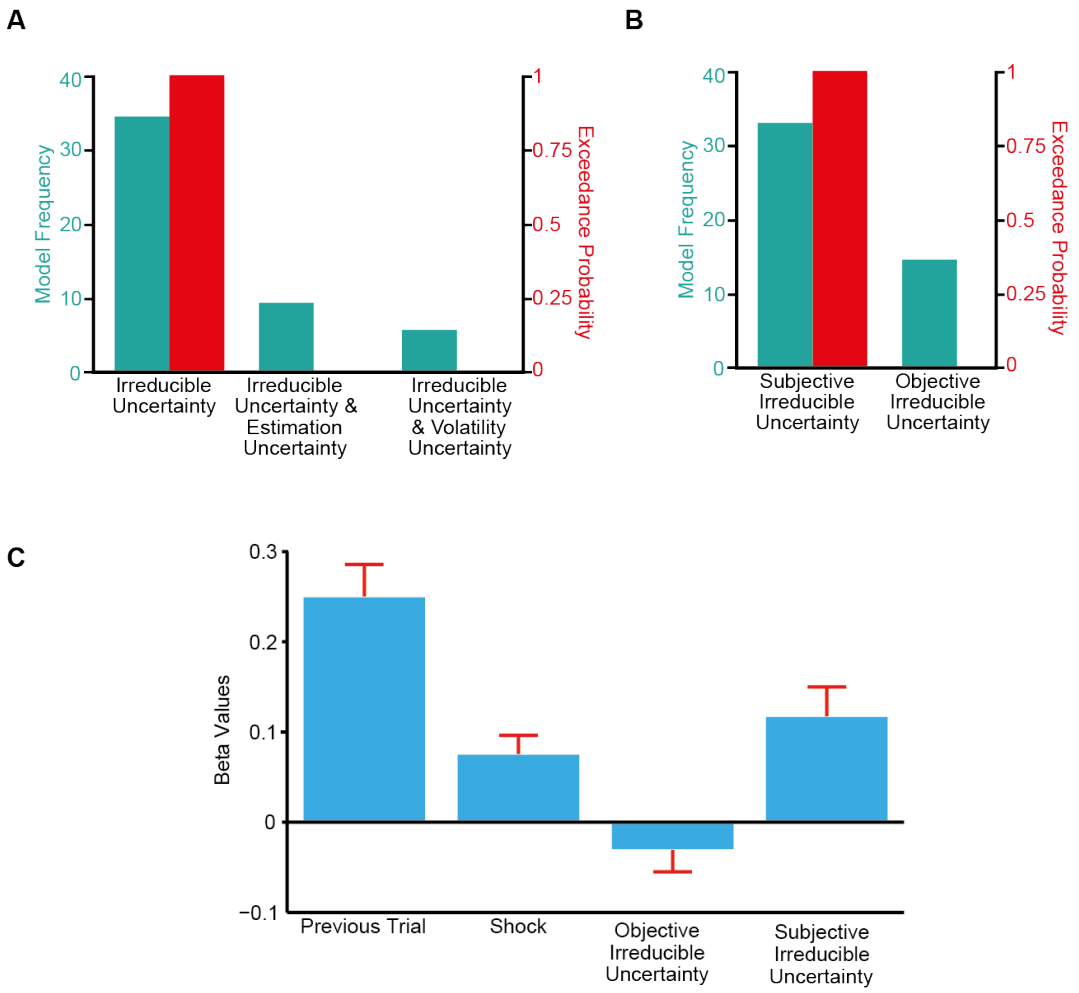
Our multiple stress measures reflect the view that stress is a multidimensional construct, expressed through subjective and physiological channels [43]. Some theoretical accounts highlight the importance of 'coherence' between stress systems in response to a challenge [29], though

discordance between the amplitude of physiological and subjective stress responses is rife (reviewed by Campbell & Ehlert [39]). This is in part because the stressors typically used in laboratory experiments with humans, such as social stress, are difficult to parameterise, precluding a detailed quantitative analysis. Conversely, using a quantitative computational approach, we show that emotional and physiological stress responses track uncertainty and are correlated within individuals. We do not claim that ours are exhaustive metrics of stress, nor that chronic stress necessarily behaves similarly to the acute stress examined here. Establishing how acute stress accumulates to produce allostasis [60], and what effects such allostasis exerts upon subsequent acute stress responses, is a major challenge for the field.

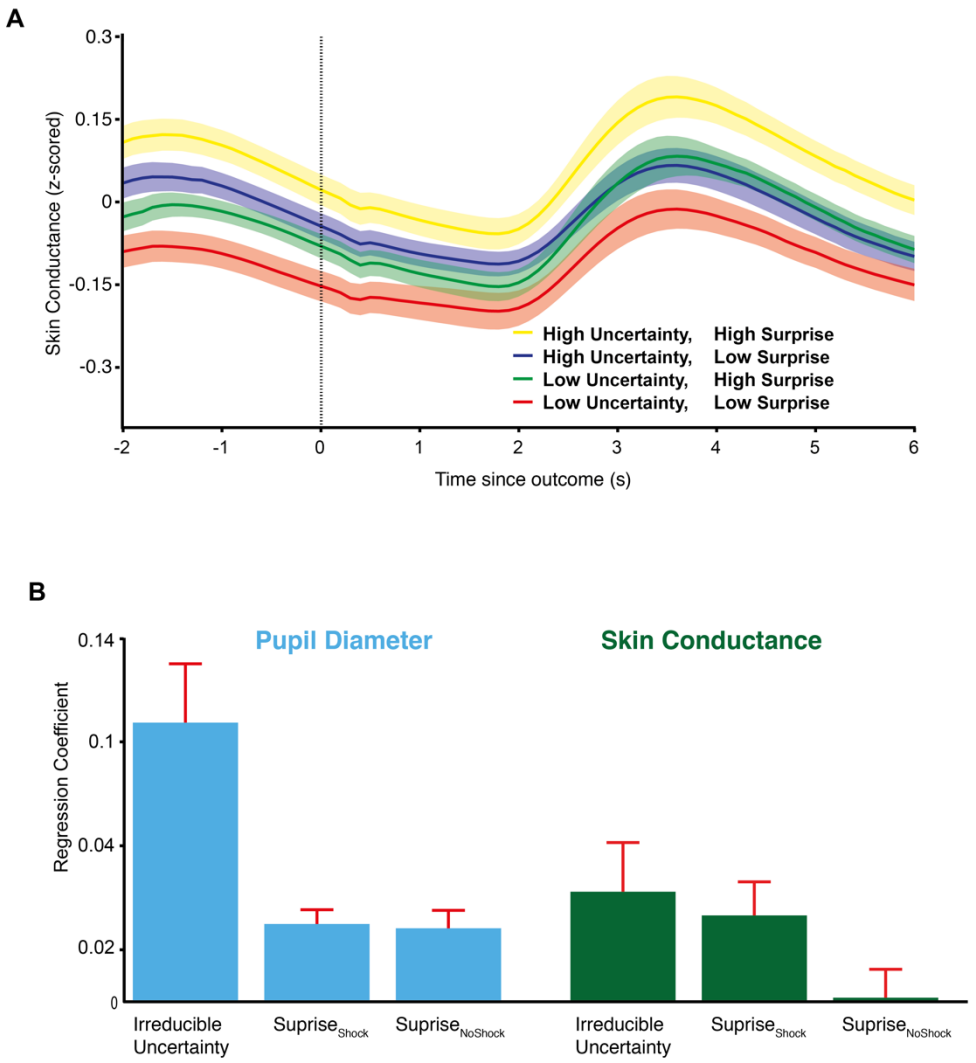
4.5.3 Conclusion

An integrated understanding of normal brain function and its perturbation in disease will require detailed analysis at multiple levels of description, from behavioural to cellular. Here we provide a computational account of acute stress responses in humans. Our results emphasise the utility of formal models of learning, particularly those in a Bayesian tradition, to understanding emotional dynamics. Dysfunction of stress response systems is common to many psychiatric disorders [2] suggesting that a computational decomposition of stress responses of the kind provide here may prove a fruitful addition to the nascent field of computational psychiatry [61].

4.6 Supplementary figures

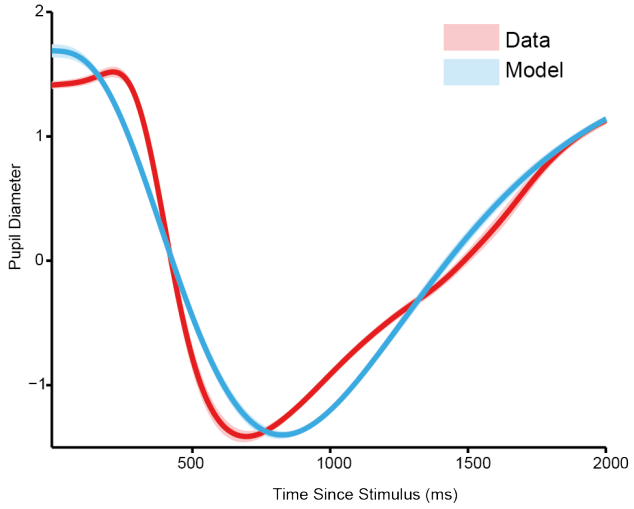


Supplementary Figure 4.1 | Assessing alternative models of subjective stress (A) Model comparison demonstrated that adding additional estimation or volatility uncertainty did not improve the performance of the irreducible uncertainty model used to explain subjective stress responses (Model Frequency=70.5%, Exceedance Probability~1). **(B)** Model comparison confirmed that subjective uncertainty as furnished by the HGF model provided a better predictor of subjective stress (Model Frequency=70.5%, Exceedance Probability=0.993) than objective uncertainty. **(C)** As expected, objective and subjective uncertainty are correlated (Pearson correlation, mean r=0.452); if both are placed in the model, only subjective uncertainty provides a significant predictor of subjective stress. Error bars are SEM across participants.

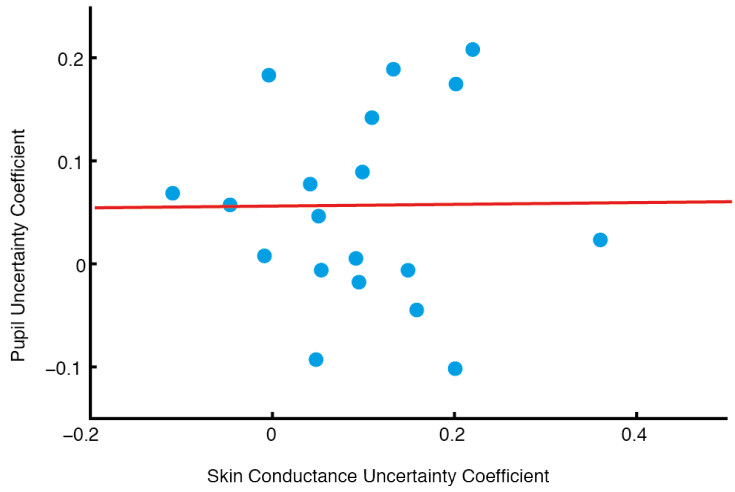


Supplementary Figure 4.2 | Additional physiological stress data (A) We split trials by uncertainty and surprise (see main results and Figure 4.4A). Average skin conductance was higher for trials with higher uncertainty and surprise (Repeated Measures ANOVA, Uncertainty: $F_{1,36}$=9.36, p=0.004, η^2=0.104, Surprise: $F_{1,36}$=12.40, p=0.001, η^2=0.070), with no evidence of an interaction ($F_{1,36}$=0.14, p=0.71). **(B)** Our regression models of pupil diameter and skin conductance included regressors for surprise at outcome,to ensure that correlations with surprise were not attributed to uncertainty. Uncertainty was a significant predictor of both pupil diameter (robust regression, β=0.11, single-sample t-test t_{21}=4.72, p<0.001) and skin conductance (robust regression, β=0.044, single-sample t-test t_{36}=2.25, p=0.031), with surprise predicting pupil diameter equally on shock and no shock trials (robust regression, $\text{Surprise}_{\text{Shock}}$: β=0.018, single-sample t-test t_{21}=4.11, p<0.001; $\text{Surprise}_{\text{NoShock}}$: β=0.017, t_{21}=3.05, p=0.0060; Difference, paired t-test t_{21}=0.209, p=0.84). However, skin conductance reflected surprise asymmetrically, with the parameter for surprise on shock trials greater than that for no-shock trials (robust regression, $\text{Surprise}_{\text{Shock}}$: β=0.035, single-sample t-test t_{36}=2.58, p=0.014; $\text{Surprise}_{\text{NoShock}}$: β=0.0019, single-sample t-test t_{36}=0.167, p=0.87; Difference, paired t-test

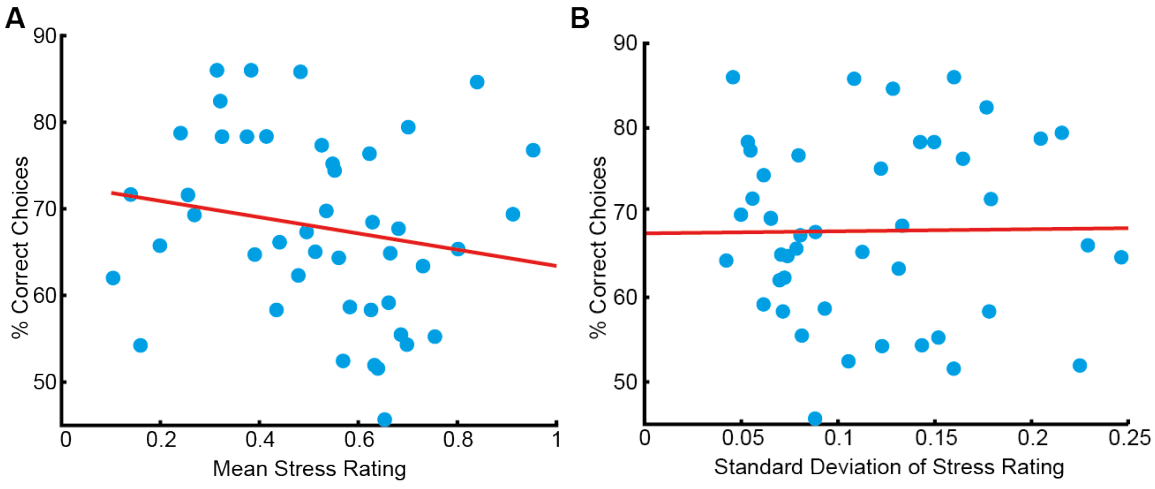
t_{36}=1.84, p=0.07). This replicates previous observations of asymmetric prediction error representations in skin conductance measurements[69]. However, a direct comparison of parameters from participants for whom we recorded both pupil diameter and skin conductance suggested that the two were not significantly different, with neither a main effect (Repeated Measures ANOVA, Effect of recording modality, $F_{1,35}$=2.82, p=0.11, η^2=0.025) nor an interaction between modality and regressor (Repeated Measures ANOVA, Modality x Regressor interaction, $F_{1,35}$=0.96, p=0.39, η^2=0.012). We are therefore unable to reject the null hypothesis that pupil diameter and skin conductance track aversive learning in comparable manners; disparities between the two may therefore be a result of low signal-to-noise in skin conductance measurements. Error bars are SEM across participants.



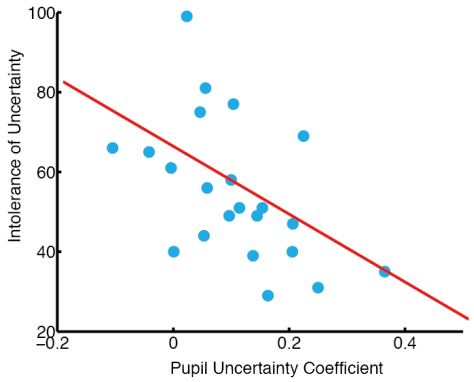
Supplementary Figure 4.3 | Luminance fitting procedure used for model of pupil diameter. In order to accommodate fluctuations in pupil diameter induced by luminance changes, we fit a luminance response function for each subject. This was conducted using a reference data set acquired after the experiment, in which each of the images used in the experiment was shown 50 times. We then fit a single luminance response function using data from all image presentations (red trace above, shading is SEM). We used the gamma response function defined in[41] which has two free parameters: time to peak (T_{max}) and number of cascade components (n). We found the best fitting pair of parameters for our luminance response function using gradient descent methods implemented by the MatLab function fmincon (blue trace above, shading is SEM). As expected from the fast constriction typically associated with the light reflex, the average T_{max} in our luminance response function was smaller than that used in the conventional response function (839ms vs 930ms). We were thus able to include luminance responses in our regression models of pupil diameter (Figure 4.5 and Supplementary Figure 4.2B).



Supplementary Figure 4.4 | Pupillary and skin conductance sensitivity to uncertainty are uncorrelated. We found no evidence of correlation between pupillary and skin conductance sensitivity to uncertainty (Pearson correlation, *n*=19, *r*=0, *p*=0.97). This is in contrast to the positive correlation observed between each of these parameters and subjective stress uncertainty sensitivity (Figure 4.5D).



Supplementary Figure 4.5 | Mean and variance of subjective stress ratings are unrelated to performance. Neither mean **(A)** (Pearson correlation, *n*=45, *r*=-0.18, *p*=0.23) nor standard deviation **(B)** (Pearson correlation *n*=45, *r*=0.014, *p*=0.93) of subjective stress ratings relate to performance on our task. Conversely, computational modelling reveals that the 'uncertainty-tuning' of stress responses based on dynamic uncertainty estimates is correlated with task performance (Figure 4.6E).



Supplementary Figure 4.6 | Uncertainty-tuning in the pupil is inversely correlated with Intolerance of Uncertainty. Pupillary sensitivity to uncertainty was inversely related to a questionnaire measure of uncertainty aversion (Intolerance of Uncertainty[70]). Subjects with greater pupil sensitivity to uncertainty were less averse to uncertainty (Pearson correlation, n=22, r=-0.51, p=0.015). This may be related to the other effect we observed, that performance under uncertainty is predicted by the sensitivity of the pupil to uncertainty (Figure 4.6B); aversion to uncertainty may be a preference rooted in the fact that individuals whose pupils do not track uncertainty perform poorly when in uncertain situations.

Supplementary Table 4.1: Parameters used in pupil model

Parameter	Mean	SEM	Tt_{21}	*p*	Form
Constant	0.0614	0.0229	2.68	0.0140	
Stimuli*	-0.0704	0.0165	-4.27	<0.001	Delta function, aligned to stimuli presentations, luminance convolved
Outcome	-0.6922	0.1774	-3.90	<0.001	Delta function, aligned to all outcomes, luminance convolved
Shocks	0.4930	0.1042	4.73	<0.001	Delta function aligned to shock outcomes
No Shocks	0.2561	0.1106	2.32	0.0308	Delta function aligned to no shock outcomes
Surprise $_{\text{Shock}}$	0.0183	0.0045	4.11	<0.001	Delta function aligned to shocks and scaled by trial surprise ($\lvert\delta_1\rvert$)
Surprise $_{\text{No Shock}}$	0.0167	0.0055	3.05	0.0060	Delta function aligned to no shocks and scaled by trial surprise ($\lvert\delta_1\rvert$)
Irreducible Uncertainty	0.1069	0.0227	4.72	<0.001	Boxcar from stimulus onset, scaled by trial irreducible uncertainty (σ_1)
Gaze X coordinate	-0.0501	0.0102	-4.90	<0.001	Unconvolved
Gaze Y coordinate	0.0051	0.0218	0.23	0.818	Unconvolved
Predictors were convolved with a standard pupillary response function (see SI Methods).The exceptions were the gaze X and Y coordinates, which were unconvolved, and the Stimuli regressor. For all phasic responses (Stimuli, Outcome, Shock, No Shock, Surprise Shock, Surprise No Shock), we also convolved predictors with first and second derivatives of the response function to allow for variance in the shape and timing of the response (data not shown).					

* The Stimuli regressor was not convolved with the canonical pupillary response function, but with the luminance response function estimated for each subject. See Supplementary Figure 3 for details.

Supplementary Table 4.2: Parameters used in skin conductance models

Parameter	Mean	SEM	Tt_{21}	*p*	Form
Constant	-0.1054	0.0111	-9.47	<0.001	
Stimuli	-0.0093	0.0087	-1.07	0.2905	Delta function, aligned to stimuli presentations, luminance convolved
Outcome	0.1177	0.0260	4.53	<0.001	Delta function, aligned to all outcomes, luminance convolved
Shocks	0.0698	0.0242	2.89	0.0066	Delta function aligned to shock outcomes
No Shocks	-0.1890	0.0347	-5.45	<0.001	Delta function aligned to no shock outcomes
Surprise Shock	0.0347	0.0134	2.59	0.0139	Delta function aligned to shocks and scaled by trial surprise (|δ1|)
Surprise No Shock	0.0019	0.0113	0.17	0.8686	Delta function aligned to no shocks and scaled by trial surprise (|δ1|)
Irreducible Uncertainty	0.0442	0.0196	2.25	0.0305	Boxcar from stimulus onset, scaled by trial irreducible uncertainty (σ1)

Predictors were convolved with a standard skin conductance response function (see SI Methods).

We also included regressors for each block (i.e. each stretch of 10 minutes between breaks) to account for changes in baseline between blocks.

As with the pupil model, for all phasic responses (Stimuli, Outcome, Shock, No Shock, Surprise Shock, Surprise No Shock), we also convolved predictors with first and second derivatives of the response function to allow for variance in the shape and timing of the response.

Supplementary Table 4.3: Hierarchical Gaussian Filter details

Parameter	Notes	Value
Model constants		
ϑ	Metavolatility parameter, controlling step size at the third level. Estimated in logit space.	Mean = 0
		Variance = 16
ω	Constant component of the learning rate at the second level. Estimated in native space.	Mean = -2
		Variance = 16
κ	Modulates coupling between 3rd and 2nd levels. Held constant.	
Trajectories Note that since uncertainty (σ) has a natural lower bound at zero – one cannot have negative uncertainty – it is estimated in log space. The numbers given here refer to values in that space.		
Predictions (X_1)	The predictions are a sigmoid transformation of the probabilities represented in X_2, and so do not have a starting value.	$\hat{\mu}_{1:}$ Mean = none Variance = none
		$\hat{\sigma}_{1:}$ Mean = none Variance = none
Probabilities (X_2)	A starting value of 0 implies neutrality between outcomes. Starting variance was chosen to be Bayes optimal using the tools provided in the TAPAS toolbox ('tapas_bayes_optimal_binary_config').	$\hat{\mu}_{2:}$ Mean = 0 Variance = 0
		$\hat{\sigma}_{2:}$ Mean = 0.06 Variance = 0
Volatility (X_3)	The absolute starting value of X_3 is arbitrary, as changes in fitted parameters will affect scaling.	$\hat{\mu}_{3:}$ Mean = 1 Variance = 0
		$\hat{\sigma}_{3:}$ Mean = 4 Variance = 0

Supplementary Table 4.4: Details of each learning model used

Model	**Notes**	**Estimated parameters: mean (standard deviation)**
Rescorla-Wagner	Beliefs are symmetrically updated, with a learning rate fitted to each subject.	α =0.38 (0.21)
Asymmetric Rescorla-Wagner	Beliefs are asymmetrically updated, with beliefs about the two stimuli updated individually.	α =0.47 (0.25)
Dual Learning Rate Rescorla-Wagner	Beliefs are updated with different learning rates on shock and no shock trials; two learning rates fitted to each subject.	α_{Shock}=0.36 (0.24) $\alpha_{NoShock}$=0.35 (0.24)
Sutton K1	Beliefs updated with a variable learning rate that depends upon the amplitude of recent prediction errors.	μ =1.65 (3.00) ν =0.53 (0.31) *h*=0.005 (0.002)
HGF	Three layer model with two fitted parameters governing connections between layers and step size at the top layer.	θ =0.034 (0.02) ω =-2.80 (2.43)

4.7 References

1. McEwen, B. S. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews* 87, 873–904 (2007).
2. de Kloet, E. R., Joëls, M. & Holsboer, F. Stress and the brain: from adaptation to disease. *Nature Reviews Neuroscience* 6, 463–475 (2005).
3. Amat, J. *et al.* Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature Neuroscience* 8, 365–371 (2005).
4. Koolhaas, J. M. *et al.* Stress revisited: a critical evaluation of the stress concept. *Neurosci Biobehav Rev* 35, 1291–1301 (2011).
5. Miller, S. M. Controllability and human stress: method, evidence and theory. *Behaviour Research and Therapy* 17, 287–304 (1979).
6. Monat, A., Averill, J. R. & Lazarus, R. S. Anticipatory stress and coping reactions under various conditions of uncertainty. *Journal of Personality and Social Psychology* 24, 237–253 (1972).
7. Pervin, L. A. The need to predict and control under conditions of threat. *Journal of Personality* 31, 570–587 (1963).
8. Weiss, J. M. Effects of coping behavior in different warning signal conditions on stress pathology in rats. *J Comp Physiol Psychol* 77, 1–13 (1971).
9. Carlsson, K. *et al.* Predictability modulates the affective and sensory-discriminative neural processing of pain. *Neuroimage* 32, 1804–1814 (2006).
10. Seidel, E.-M. *et al.* Uncertainty during pain anticipation: The adaptive value of preparatory processes. *Hum Brain Mapp* 36, 744–755 (2014).
11. Yoshida, W., Seymour, B., Koltzenburg, M. & Dolan, R. J. Uncertainty increases pain: evidence for a novel mechanism of pain modulation involving the periaqueductal gray. *J. Neurosci.* 33, 5638–5646 (2013).
12. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692 (2005).
13. Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S. & Krugers, H. J. Learning under stress: how does it work? *Trends Cogn. Sci. (Regul. Ed.)* 10, 152–158 (2006).
14. Russo, S. J., Murrough, J. W., Han, M.-H., Charney, D. S. & Nestler, E. J. Neurobiology of resilience. *Nature Neuroscience* 15, 1475–1484 (2012).
15. Iglesias, S. *et al.* Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* 80, 519–530 (2013).
16. Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience* 14, 488–501 (2013).
17. Payzan-LeNestour, E. & Bossaerts, P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comp Biol* 7, e1001048 (2011).
18. Bland, A. R. & Schaefer, A. Different varieties of uncertainty in human decision-making. *Frontiers in Neuroscience* 6, 85 (2012).
19. Bach, D. R. & Dolan, R. J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience* 13, 572–586 (2012).
20. Payzan-LeNestour, E., Dunne, S., Bossaerts, P. & O'Doherty, J. P. The neural representation of unexpected uncertainty during value-based decision making. *Neuron* 79, 191–201 (2013).
21. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 5, 39 (2011).
22. Keinan, G. Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats. *Journal of Personality and Social Psychology* 52, 639–644 (1987).
23. Robinson, O. J., Overstreet, C., Charney, D. R., Vytal, K. & Grillon, C. Stress increases aversive prediction error signal in the ventral striatum. *Proceedings of the National Academy of Sciences* 110, 4129–4133 (2013).
24. Averill, J. R. & Rosenn, M. Vigilant and nonvigilant coping strategies and psychophysiological stress reactions during the anticipation of electric shock. *Journal of Personality and Social Psychology* 23,

128–141 (1972).
25. Bali, A. & Jaggi, A. S. Preclinical experimental stress studies: Protocols, assessment and comparison. *Eur. J. Pharmacol.* 746, 282–292 (2015).
26. Hellhammer, D. H., Wüst, S. & Kudielka, B. M. Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology* 34, 163–171 (2009).
27. Hermans, E. J. *et al.* Stress-related noradrenergic activity prompts large-scale neural network reconfiguration. *Science* 334, 1151–1153 (2011).
28. Hermans, E. J., Henckens, M. J. A. G., Joëls, M. & Fernández, G. Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in neurosciences* (2014). doi:10.1016/j.tins.2014.03.006
29. Andrews, J., Ali, N. & Pruessner, J. C. Reflections on the interaction of psychogenic stress systems in humans: The stress coherence/compensation model. *Psychoneuroendocrinology* 38, 947–961 (2013).
30. Henckens, M. J. A. G., Hermans, E. J., Pu, Z., Joëls, M. & Fernández, G. Stressed memories: how acute stress affects memory formation in humans. *J. Neurosci.* 29, 10111–10119 (2009).
31. Bradley, M. M., Miccoli, L., Escrig, M. A. & Lang, P. J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 602–607 (2008).
32. Baker, S. R. & Stephenson, D. Prediction and control as determinants of behavioural uncertainty: effects on task performance and heart rate reactivity. *Integr Physiol Behav Sci* 35, 235–250 (2000).
33. Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences* 110, 20941–20946 (2013).
34. Rescorla, R. A. & Wagner, A. R. in *Classical conditioning: current research and theory* (eds. Black, A. H. & Prokasy, W. F.) (Appleton-Century-Crofts, New York Meredith Division, 1972).
35. Sutton, R. S. Gain adaptation beats least squares. in (1992).
36. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* 46, 1004–1017 (2009).
37. Daunizeau, J., Adam, V. & Rigoux, L. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Comp Biol* 10, e1003441 (2014).
38. Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress. *J Health Soc Behav* 24, 385–396 (1983).
39. Campbell, J. & Ehlert, U. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology* 37, 1111–1134 (2012).
40. Spoormaker, V. I. *et al.* Additional support for the existence of skin conductance responses at unconditioned stimulus omission. *Neuroimage* 63, 1404–1407 (2012).
41. de Gee, J. W., Knapen, T. & Donner, T. H. Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences* 111, E618–25 (2014).
42. Bach, D. R., Flandin, G., Friston, K. J. & Dolan, R. J. Modelling event-related skin conductance responses. *International Journal of Psychophysiology* 75, 349–356 (2010).
43. Scherer, K. R. What are emotions? And how can they be measured? *Social Science Information* 44, 695–729 (2005).
44. Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H. & Gross, J. J. The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion* 5, 175–190 (2005).
45. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences* 111, 12252–12257 (2014).
46. Lukšys, G., Gerstner, W. & Sandi, C. Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning. *Nature Neuroscience* 12, 1180–1186 (2009).
47. Shou, W., Bergstrom, C. T., Chakraborty, A. K. & Skinner, F. K. Theory, models and biology. *eLife Sciences* 4, (2015).
48. Stephan, K. E., Iglesias, S., Heinzle, J. & Diaconescu, A. O. Translational Perspectives for

Computational Neuroimaging. *Neuron* 87, 716–732 (2015).
49. Montague, P. R., Dolan, R. J., Friston, K. J. & Dayan, P. Computational psychiatry. *Trends Cogn. Sci. (Regul. Ed.)* 16, 72–80 (2012).
50. Schultz, W., Dayan, P. & al, E. A neural substrate of prediction and reward. *Science* (1997).
51. Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H. & Balsters, J. H. Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum Brain Mapp* 35, 4140–4154 (2014).
52. Joshi, S., Li, Y., Kalwani, R. M. & Gold, J. I. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron* 89, 221–234 (2016).
53. Ulrich-Lai, Y. M. & Herman, J. P. Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience* 10, 397–409 (2009).
54. Schwabe, L., Tegenthoff, M., Hoffken, O. & Wolf, O. T. Concurrent glucocorticoid and noradrenergic activity shifts instrumental behavior from goal-directed to habitual control. *J. Neurosci.* 30, 8190–8196 (2010).
55. Strange, B. A. & Dolan, R. β-Adrenergic modulation of emotional memory-evoked human amygdala and hippocampal responses. *Proceedings of the National Academy of Sciences* 101, 11454–11458 (2004).
56. Preuschoff, K., 't Hart, B. M. & Einhäuser, W. Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience* 5, 115 (2011).
57. Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience* 15, 1040–1046 (2012).
58. Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X. & Bishop, S. J. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience* 18, 590–596 (2015).
59. Tomb, I., Hauser, M., Deldin, P. & Caramazza, A. Do somatic markers mediate decisions on the gambling task? *Nature Neuroscience* 5, 1103–1104 (2002).
60. McEwen, B. S. & Gianaros, P. J. Stress- and allostasis-induced brain plasticity. *Annu. Rev. Med.* 62, 431–445 (2011).
61. Wang, X.-J. & Krystal, J. H. Computational Psychiatry. *Neuron* 84, 638–654 (2014).
62. Gracely, R. H., Lota, L., Walter, D. J. & Dubner, R. A multiple random staircase method of psychophysical pain assessment. *Pain* 32, 55–63 (1988).
63. Watson, A. B. & Pelli, D. G. Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics* 33, 113–120 (1983).
64. Arakawa, H., Maeda, M. & Tsuji, A. Chemiluminescence enzyme immunoassay of cortisol using peroxidase as label. *Analytical Biochemistry* 97, 248–254 (1979).
65. Friston, K. J. *et al.* Event-related fMRI: characterizing differential responses. *Neuroimage* 7, 30–40 (1998).
66. Schwarz, G. Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461–464 (1978).
67. Bach, D. R., Flandin, G., Friston, K. J. & Dolan, R. J. Time-series analysis for rapid event-related skin conductance responses. *J. Neurosci. Methods* 184, 224–234 (2009).
68. Hoeks, B. & Levelt, W. J. M. Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods, Instruments, & Computers* 25, 16–26 (1993).
69. Bach, D. R. & Friston, K. J. No evidence for a negative prediction error signal in peripheral indicators of sympathetic arousal. *Neuroimage* 59, 883–884 (2012).
70. Buhr, K. & Dugas, M. J. The Intolerance of Uncertainty Scale: psychometric properties of the English version. *Behaviour Research and Therapy* **40,** 931–945 (2002).

Chapter 5: Formation of value from quality and quantity in human decision making

5.1 Abstract

In the previous chapters we have seen how learning models can be applied to understand the generation and impact of emotion. In **Chapters 2** and **3**, we used reinforcement learning models to represent how organisms learn about the value of stimuli through repeated experience. In many cases, however, value estimates are computed 'on the fly', on the basis of multiple pieces of data. In this chapter we explore the neural process underlying a common case of this process, in which information about quality and quantity must be combined to assess the value of an option. We used giftcards, allowing us to vary the quality (by switching between stores, which have distinct values to our participants) and the quantity (by changing the amount of money on each card). Using fMRI, we examined correlates of quality, quantity, and their integration into value signals. We found that the Inferior Frontal Gyrus (IFG) correlated with the giftcard quality, whilst the Intra Parietal Sulcus (IPS) specifically reflected quantity. Several brain regions, including the cingulate cortex, were sensitive to the interaction of quality and quantity, implying a role in the computation of value. We found that the Anterior Cingulate Cortex (ACC) was uniquely activated by quality, quantity, and their interaction, suggesting a hub for the flexible calculation of value from quality and quantity. Additional Repetition Suppression (RS) analyses identified a region immediately posterior to the value-integrating portion of the ACC which displayed patterns of activity consistent with a tuned representation of value.

5.2 Introduction

As reviewed in section 1.2, several decades of work on value-based decision-making have identified numerous correlates of value in the brain [1]. Convergent evidence from human fMRI [2-8] and non-human primate recordings [9-14] suggest that animals form neural representations of the subjective value of stimuli in a wide variety of brain areas, potentially in an automatic fashion invariant to the task at hand [15-16].

Value representations might be useful for a wide variety of computations regularly performed by the brain, such as the allocation of attention [17], energy [18], and mnemonic resources [19]. Most prominently, these value estimates are thought to form the input to value-comparison mechanisms whereby decision are made between options [21], by a process variously characterized as evidence accumulation [21,22] or mutually inhibitory competition [23-26]. Representations of stimulus value also play a crucial role in the models of reinforcement learning discussed earlier in this thesis, in which discrepancies between experienced and expected values result in prediction errors that drive learning [30].

5.2.1 How are value estimates constructed?

Despite abundant evidence that value representations exist, we know little about how they come about. Efforts to isolate value signals in a neuroeconomic framework have involved carefully controlling stimulus characteristics and action requirements to distill value from its component parts[31,32]. However, recent studies have emphasized that biological and evolutionary contextualization of decision-making provides a richer account of the representations and computations which underlie choice [33-35]. An appreciation of the evolutionary pressures faced by an agent allows us to ask how the brain represents variables relevant to survival, rather than those considered important by economic theory.

One fruitful avenue of exploration has been the use of foraging tasks [36-38], in which animals make a series of decisions about whether to sample a currently presented option, or move on in the hope of obtaining something better. Such tasks provide insight into the encoding of ethologically relevant variables, like the value of searching the environment for alternative options, in addition to more traditional analyses of option value [9,39,40]. In this experiment we

drew inspiration from foraging tasks to ask how current option value is constructed from its two component parts: quality and quantity.

Quality (how good a specific reward is) and quantity (how much of it there is) arguably represent the cardinal dimensions from which the value of rewarding stimuli is calculated. Previous studies typically vary one or both to manipulate value, but to date a characterization of their independent representation and combination is lacking. The calculation of value from quality and quantity is a basic model-based computation (see section 1.3.2.4), requiring animals to combine estimates of quality with information about quantity*. Many primate studies vary both the number of drops and the subjective quality of different juices, but focus on the resultant menu-independent and persistent value representation [9,39,40]. Conversely, human neuroimaging studies often involve choices between goods that are matched in quantity, but vary in quality (i.e. packet of crisps vs. chocolate bar) [2,41-44], or rely upon financial rewards that vary only in quantity [45-47]. In both cases, how quality and quantity are processed in parallel before being combined to calculate an option value remains unclear.

We designed an experiment in which participants integrated information about the quality of a giftcard (how subjectively valuable it was for them to be able to spend money at a particular store) and quantity (how many £ were on the giftcard). In a behavioural session we characterized the combination of quality and quantity into integrated value using an auction procedure (Becker-DeGroot-Marschak, BDM) [48], which allowed us to select giftcards with distinct qualities for use in an fMRI experiment. In the MRI scanner, participants evaluated a series of giftcards without choice, allowing us to examine the representation of quality, quantity, and value, free from decision-related signals [49]. Interspersed choices ensured engagement and encouraged participants to evaluate displayed options.

We asked two questions about value representations: which brain regions participate in the *computation* of value from its component parts, and what is the *form* of the resultant representation? To address the first question, we characterized the relationship between quality,

---

* Quality estimates might be obtained either in a model-based fashion or a model-free fashion. However, the combination of quality and quantity information requires a basic model of the world, in the form of knowledge about how the value of an offer scales with quantity.

quantity, and value, and looked for correlates of each in the brain: areas in which activity was higher when the variable of interest was greater. However, such analyses make an assumption about the underlying neural processes, namely that neurons fire more (or less) in order to encode the value of a single variable. This assumption dominates both fMRI and single-cell analyses of value. Bar a few recent studies using classification analyses [42-44], analysis of value representation is overwhelmingly conducted using General Linear Models (GLMs), sensitive to linear relationships between the dependent variable (BOLD activity in a voxel or firing rate in a neuron) and the independent variable (value).

5.2.2 How are value estimates represented?

There are (at least) three reasons to question the assumption that value is coded linearly. Firstly, recent analyses of prefrontal cortex (PFC) activity have uncovered rich, flexible, and temporally diverse codes which suggest that simple monotonic coding is a poor model of tuning curves in the PFC [50-52]. Given that the prefrontal and cingulate cortices are a hub for value-guided decision-making [53], these studies suggest that representations of value might be more sophisticated than previously appreciated. Secondly, value estimates must be highly malleable, preferably incorporating some measure of uncertainty in order to allow appropriate updating [54,55], as reviewed in the introduction section 1.4.3. Extant work has focused upon the segregated encoding of value-uncertainty by neurons distinct from the ones carrying value-magnitude information, thus assuming a 'summary-statistics' representation [56-63]. Conversely, theoretical treatments of sensory coding suggest that uncertainty could be folded into the representation itself, elegantly conferring the ability to perform Bayesian integration via summation [64,65]. As described in section 1.4.6, a prerequisite for such coding is that the variable in question is coded by a population of non-linearly tuned neurons. The fact that just such a non-linear encoding scheme is found in the representation of number in parietal and frontal cortices [66-68] provides a third reason to think that value coding might be non-linear, or even Gaussian-tuned, in some parts of the brain.

We therefore constructed our experiment to make use of Repetition Suppression (RS). RS describes the phenomenon whereby repeated presentation of the same or similar stimuli elicits a reduction in evoked responses. By examining the influence of the previous trial upon activity in the current trial, RS can reveal structure to neural representations that is invisible to traditional,

parametric approaches [69]. Although the cellular mechanisms of RS are unclear and probably diverse [70,71], numerous studies confirm that the amount of repetition suppression exerts by one stimulus upon another is modulated by the overlap in their neural representation, with maximal suppression elicited by repetition of the same stimulus [72-77]. RS has recently been used to elucidate representations of outcomes and agents in value-based decision-making [78-80], the neural representation of space [81], and the organization of internal models of stimuli and their associations [82,83]. Particularly salient to the current experiment, RS has provided evidence for Gaussian tuning for numerosity in humans [84-86], replicating recordings in non-human primates [67].

We therefore sought to use RS as an assay for non-linear tuning for value, as predicted by the considerations in the previous section. Concretely, we hypothesized that value might be represented by populations of tuned neurons with overlapping, non-linear tuning for value, as previously documented for number [67]. Following the logic of previous repetition studies, presentation of a stimulus of a given value elicits repetition suppression in those neurons sensitive to the presented value. This means that subsequent stimuli which activate the same, or an overlapping, population of neurons, elicit less activity than stimuli activating un-suppressed populations of neurons. The aggregate effect, visible at the voxel-level, is that activity evoked on a trial is a function of how dissimilar the current stimulus is to the previous one [84,85]. This idea is explored more in due course – see Figure 5.8 for a graphical description.

To preview our results, we found that quality was represented in the Inferior Frontal Gyrus (IFG), extending into the lateral PFC. Conversely, quantity was associated with increasing activity in the bilateral Intraparietal Sulcus (IPS). The interaction between the two (higher slope of quantity coding with higher quality) correlated with activity in the posterior cingulate cortex and bilaterally in the superior temporal lobe. We found that an Anterior Cingulate Cortex (ACC) displayed a conjunction of all three effects, suggesting that it provides a focal point for the calculation of integrated value from its component parts. Consistent with this, we observed repetition suppression for integrated value in the cingulate; activity covaried with the absolute difference in value between consecutive trials. However, further analysis of a neural firing-rate model suggests that such signals might arise from mixed linear codes, and not Gaussian-tuning as previously supposed [84].

5.3 Methods

5.3.1 Participants

47 participants participated in the behavioural study, with 26 returning for the fMRI session. Of these, one participant failed to complete the experiment due to ill-health, leaving 25 participants in the imaging study. Both studies were approved by a local ethics committee (Research Ethics Committee UCL, ref. 3450/002).

5.3.2 Stimuli & payment

Based on pilot experiments, we selected 13 giftcards according to several criteria. We chose cards that were well known to the participant population, maximised between-subject variability, and displayed minimal correlations between cards (i.e. preferences for a given card were hard to predict from preferences for other cards).

During the behavioural session, participants completed two tasks: an auction procedure, from which they could obtain a mixture of up to £20 cash and a £20 giftcard, and a session of paired choices between cards worth £20 (Figure 5.1). One trial was randomly selected across both sessions and reimbursed appropriately.

For the fMRI experiment, participants first completed paired choices between cards worth £20 outside of the scanner, and subsequently chose between cards worth £1-20 within the scanner. One trial from each task was reimbursed, in addition to a £20 flat rate for experiment completion.

5.3.3 Behavioural session

Participants first performed an auction task (Becker-DeGroot-Marshak, BDM) designed to elicit their subjective valuation of different giftcards holding varying amounts of money [48]. Briefly, the BDM involves players placing a minimum bid for an item on each trial. After the experiment, a single trial is randomly selected for reimbursement. For that trial, a randomly drawn number – the 'cost'- is compared to the bid. If the cost is higher than the bid, the player retains their endowment and does not receive the item. If the bid is higher than the cost, then the player receives the item, but, crucially, pays the cost rather than their bid. This removes the incentive to place low bids, resulting in an optimal strategy in which players report their true values.

Prior to the task, participants were led through an example trial and answered questions confirming understanding. Each of 13 giftcards was presented in association with 12 different quantities, giving a total of 156 trials.

Following the auction task, participants chose between pairs of giftcards of matched quantity (£20). Each combination of cards was presented twice, yielding 325 trials.

5.3.4 Participant and card selection for scanning study

We selected a subset of participants to complete the scanning part of the study (Figure 5.2). Selection was based upon reliability, stability, and diversity of preferences over giftcards. We fit linear regressions to values reported during the auction procedure, yielding the following equation for each giftcard:

$$\text{Integrated Value} = \beta * \text{Quantity} + C$$

Equation 5.1

Where β and C (an intercept term) were fit using robust regression (Figure 5.2A). The β's thus obtained are a measure of a giftcard's quality; the value of a single unit of currency on that giftcard. We next assessed how well these β's predicted paired choice (Figure 5.2B), selecting subjects for whom there was a close relationship (e.g. 5.2B iii).

For the scanning session, we selected three giftcards, chosen to maximise variance in quality (Figure 5.2C). We thus selected the lowest and highest quality card (max β and min β) and one closest to the mean of the two. Having performed this selection procedure, we verified that choices of these cards in the paired-choice session reflected the rankings calculated from the BDM (Figure 5.2D). These βs were used as indicators of quality for the fMRI analyses in which parametric modulators were used (GLM2 & GLM3, see below).

5.3.5 fMRI task

The task was designed to allow us to examine representations of quality, quantity, and their integration, using both linear analyses and measures of repetition suppression. To avoid measurements being confounded by variables related to the dynamics of stimulus comparison [49], on the majority of trials we presented a single stimulus (Figure 5.1D), and asked participants to evaluate its desirability.

Presentation side was flipped every 10 trials. Stimuli remained onscreen for 4000ms, before being followed by an ITI (normally distributed around 1500ms) or, in 1/7 trials, the appearance of a second giftcard. Participants were asked to make a choice between the two within 4000ms, using a button box. Failure to register a choice within this time period resulted in a 'TIME OUT' message, and participants were informed prior to scanning that if a timed-out trial was selected for reimbursement, they would receive no payment for that part of the experiment. Each giftcard displayed in the scanner was pseudo-coloured red or blue to reduce gross visual differences between cards.

Repetition suppression in fMRI effects can be sensitive to expectation [87], necessitating counterbalancing of stimulus order. We defined 7 trial types (red and blue versions of each three giftcards, + decision trials), and used a genetic algorithm to find a stimulus order in which $p(stim^i_2 | stim^j_1)$ was matched for all stimuli i and j. We manually removed trials on which decisions were repeated, leaving a sequence of 97 stimuli. The quantity (1-20) on the giftcards were randomised, effectively orthogonalising quality and quantity (mean correlation coefficient across participants = 0.0074, p= 0.40). Participants completed 4 runs of the task, yielding a total of 340 stimulus evaluation trials and 48 decision trials.

5.3.6 Logistic regression modelling of choices during fMRI task

We used logistic regression to quantify the factors modulating choices in the scanner. For each participant, we fit a model to predict whether they chose the new card (presented during the decision trial), or the old card (on-screen from the valuation trial):

$$Choice(t) = s(\beta_0 + \beta_1 Quality_{New\text{-}Old} + \beta_2 Quantity_{New\text{-}Old} + \beta_3 Interaction_{New\text{-}Old})$$

Equation 5.2

Where β_0 is a constant term accounting for option-independent biases in choice, $\beta_{1\text{-}3}$ are regression coefficients describing the effect of each term on choice, and s is the sigmoid function:

$$s(x) = \frac{1}{1 + e^{-x}}$$

Equation 5.3

Quality was defined using the betas from the BDM auction (see above and Figure 5.4A), whilst quantity merely reflected the number of £ on each giftcard. The interaction term was calculated as the mean-centred product of quality and quantity.

To assess choice predictability, we took the output of the model (valued between 0 and 1), rounded it (such that choices were either a 0 or a 1), and compared it to the vector of actual choices made by the participant. Predictability was then defined simply as the % of choices correctly predicted by the model.

5.3.7 fMRI data acquisition

Data were acquired using a Siemens 3T Trio scanner with a 32-channel head coil at the Wellcome Trust Centre for Neuroimaging.

We used a 2D Echo Planar Image (EPI) sequence optimised to minimise dropout in the OFC [88], with voxels 3mm isotropic (TR=3.36s,TE=30ms), with 48 slices giving whole brain coverage. Slices were tilted at -30^0. Scans were preceded by a field map (TE1=10ms, TE2= 12.46ms). The first 5 volumes of each run were discarded to allow for T1 equilibration.

We also acquired a T1-weighted structural scan for each subject, comprising 176 slices over a field of view of 256mm with a 1mm isotropic resolution (TR=7.92ms, TE=2.48ms) [89].

Throughout scanning, we monitored breathing rate using a pneumatic belt and pulse & blood oxygenation using an infrared pulse oximeter (Nonin systems, Model 8600 F0). Both were digitized and recorded via Spike2 (v6.17), and subsequently included in GLM analyses of brain activity along with regressors derived from motion correction [90].

5.3.8 Pre-processing

All pre-processing and data analysis took place in SPM 12 (http://www.fil.ion.ucl.ac.uk/spm/). Subsequent data visualization took place in MRICron (http://people.cas.sc.edu/rorden/mricron/index.html) and MRIcroGL (http://www.cabiatl.com/mricrogl/).

Having discarding the first 5 volumes, we corrected EPIs for field inhomogeneities using acquired field maps, bias corrected, slice-time corrected (to the middle slice), and realigned and unwarped

to the first EPI for each participant. EPIs were then co-registered to each participant's structural scan.

We used the DARTEL toolbox for between-subject registration and normalization [91]. Structural images were first segmented into white matter, grey matter, and CSF components. Segmented images were then iteratively warped into normalized MNI space, providing a template which was then used to normalize EPIs, a step which included Gaussian smoothing at 8mm FWHM.

5.3.9 fMRI data analysis

Data were analysed using a series of General Linear Models (GLMs). These were estimated for each participant, including the calculation of contrasts between different regressors (first-level analysis). This provided summary-statistics (βs) which could be tested at a population level versus a null hypothesis that they were on average equal to zero (second-level analysis) [92].

In order to obviate multiple comparisons when performing whole brain analyses, we used cluster-based correction using a cluster-defining threshold of $p<0.005$, and a cluster-corrected FWE threshold of $p<0.05$, except for in analysis of repetition suppression (GLM2, below), where a more lenient cluster-forming threshold of $p<0.01$ was used, in line with recent repetition suppression studies [78,80,82].

To extract the parameter estimates displayed in Figure 5.6, we used group-functional ROIs thresholded at $p<0.005$. For the conjunction analysis described in Figure 5.7 we took the product of 3 binary masks (quality, quantity, interaction), each thresholded at $p_{uncorrected}<0.05$, resulting in a family-wise error rate of $p_{uncorrected}=0.000125$.

5.3.9.1 GLM1: quality and quantity

Our first GLM incorporated separate onset regressors for cards of different qualities (low, medium, high). Each of these was modelled as a 4s long boxcar, and associated with a parametric modulator corresponding to the quantity on the card at each presentation. We used a fourth onset regressor corresponding to decision-trials, which were modelled as delta functions. This GLM was used to perform a whole-brain analysis of value computations during evaluation trials.

We performed three key contrasts:

Quality: $[Quality_{High} - Quality_{Low}]$

Quantity: $[Quantity_{HighQuality} + Quantity_{MediumQuality} + Quantity_{LowQuality}]$

Interaction: $[Quantity_{HighQuality} - Quantity_{LowQuality}]$

The interaction analysis was constructed to test for regions displaying steeper coding of quantity for high quality cards than low quality cards, consistent with value integration (Fig 5.2C) and corresponding to the intuition that an extra unit of a more desirable good (e.g. a Ferrari) is worth more than an extra unit of a less desirable good (e.g. an apple).

We excluded trials preceding decisions from the evaluation regressors in order to guard against contamination of the evaluation regressors by decision-related activity, a possibility due to the lack of ITI between evaluation and decision trials.

5.3.9.2 GLM2: Repetition Suppression

Following examples from the numerosity-coding literature [84-86], we designed a repetition-suppression analysis based upon the absolute change in value between trials (see Figure 8A). We used this analysis to interrogate repetition suppression effects within ROIs identified by the whole-brain analysis using GLM1.

$$\Delta IntegratedValue(t) = |IntegratedValue(t) - IntegratedValue(t\text{-}1)|$$

Equation 5.4

Where IntegratedValue(*t*) is simply the product of quality and quantity on trial *t* (as in Equation 5.1). We used a single onset regressor to represent all giftcard presentations, again using a 4s boxcar, with parametric modulators for ΔIntegrated Value, and, as a precaution, Integrated Value. The inclusion of Integrated Value in the model allowed us to confirm that any effects of ΔIntegrated Value were not the result of spurious correlation with Integated Value itself. Trials following decisions were excluded as they were preceded by a pair of stimuli, obfuscating calculation of stimulus similarity. As before, we used a second onset regressor for decision trials. Contrasts were calculated merely as the value of the relevant parametric modulators.

5.3.9.3 GLM3: Integrated Value

In order to obtain a measure of integrated value coding, we used a single 4s boxcar for all evaluation trials, associated with a parametric modulator for integrated value (Quality x Quantity), excluding pre-decision trials as in GLM1. As in GLM1 & 2, decision trials were modelled in a separate regressor with delta onsets. We used this analysis within ROIs identified by the whole-brain analysis in GLM1, to confirm that the ACC region showing a conjunction of quality, quantity, and interaction effects could also be described as coding integrated value (Figure 5.7A).

5.3.9.4 Statistical tests

Parameter estimates from fMRI are normally distributed, permitting the use of parametric statistics (t-tests and Pearson correlations). When analysing distributions we knew a priori to be non-normal (e.g. predictability, which is bounded at 0 and 100), we used non-parametric equivalents (sign-tests and Spearman rank coefficients).

All statistical testing was carried out in Matlab.

5.3.10 Neural modelling

In order to better understand our fMRI data, we constructed a simple firing rate model in which 1000 neurons were parameterised either as Gaussian or linear coders, and their response to a series of cues simulated under subtractive and divisive adaptation. We initialized each parameterisation 100 times and measured the model's response to the same 500 cues, which were randomly drawn from a uniform distribution of 1-30.

5.3.10.1 Gaussian tuning curves

For each neuron we defined a tuning curve, which was formulated as a Gaussian:

$$FR(x) = \gamma e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Equation 5.5

Where γ is a scaling parameter, μ is the stimulus evoking the highest firing rate, randomly selected from the interval -15:45, and σ is the variance, fixed at 5. The range of μ was set so that, on average, all stimuli in the range 1:30 evoked the same amount of activity; expanding the

range of μ used thus avoided end effects. γ was set to 300, to bring the evoked firing rates into the range produced by the linear model (see below).

5.3.10.2 Linear tuning curves

Linear tuning curves were defined as:

$$FR(x) = \alpha x + \beta$$

Equation 5.6

Where α is the slope and β the intercept (equivalent to the minimum firing rate). Both slopes and intercepts were constrained to be positive, set by taking the absolute value of draws from a normal distribution centred on 0. We then 'flipped' exactly half of the neurons, to produce a mixed population of positive and negative coders, consistent with single-cell recordings in ACC [19,23,67,93].

Firing rates were then normalized such that the maximum firing rate was 30Hz.

5.3.10.3 Calculating population responses

To calculate population response to a series of cues, we calculated the activity of each neuron in the population by indexing its tuning function with the value of the current cue. We then attenuated this activation by a fraction of the activity in that neuron on the previous trial, consistent with a fatigue model of repetition suppression [70], and adding a small amount of log-normally distributed noise (mean=.05, variance=.5).

In the subtractive regime, we subtracted the adaptation term:

$$Activation(x,t) = FR(x) - \lambda * Activation(t-1) + noise$$

Equation 5.7

Where λ is an adaptation coefficient set to 0.1.

In the divisive regime, we divided by the adaptation term:

$$Activation(x,t) = FR(x) * (1 - \lambda * Activation(t-1)) + noise$$

Equation 5.8

With λ again set to 0.1.

In both cases, negative firing rates were equated to zero. We then took the mean across neurons to estimate the population response to a given cue.

5.4 Results

5.4.1 Behavioural session establishes stable quality estimates

We used a behavioural session to identify participants for whom we could find giftcards which had consistently different subjective qualities (Figure 5.1B). The behavioural session consisted of two tasks. Participants (n=45) performed a Becker-DeGroot-Marschak (BDM) auction [48] and a series of paired choices, each involving a selection of 13 giftcards (Figure 5.1C). In the BDM, players reported how much they'd be willing to pay for a giftcard loaded with a certain amount of money, from £1-20. Subsequently, participants made paired choices between different giftcards containing matched sums (£20).

We used a linear fit to the relationship between amount of money on the giftcard and amount bid for each giftcard during the BDM to provide a measure of the quality of each giftcard for each participant (Fig 5.2A). To maximize power in the fMRI study, we selected subjects whose bids were predictable (Fig 5.2A) and for whom we could select 3 giftcards with distinct qualities (Fig 5.2C, circled points in Fig 5.2B).

By way of confirmation that BDM-estimated values predicted choice, we next compared quality estimates from the BDM with the number of choices of each giftcard in the paired choice session, preferring subjects for whom there was a high correlation (Fig 5.2B).

Selected subjects thus displayed consistent BDM bids, a high correlation between preferences elicited in the BDM and paired choice sessions, and a low maximum correlation between quality, quantity, and integrated value (Figure 5.3). Integrated value was calculated as the product of quality ad quantity, effectively providing a prediction of the bid a participant would place for a given giftcard. The correlation between quality, quantity, and integrated value reflects the diversity of giftcard qualities. Giftcards with disparate qualities limit the correlation between quantity and integrated value (e.g. Fig 5.2C, row i and iii), whilst if all giftcards have similar qualities, the quantity/integrated value correlation will be high (Fig 5.2C, ii).

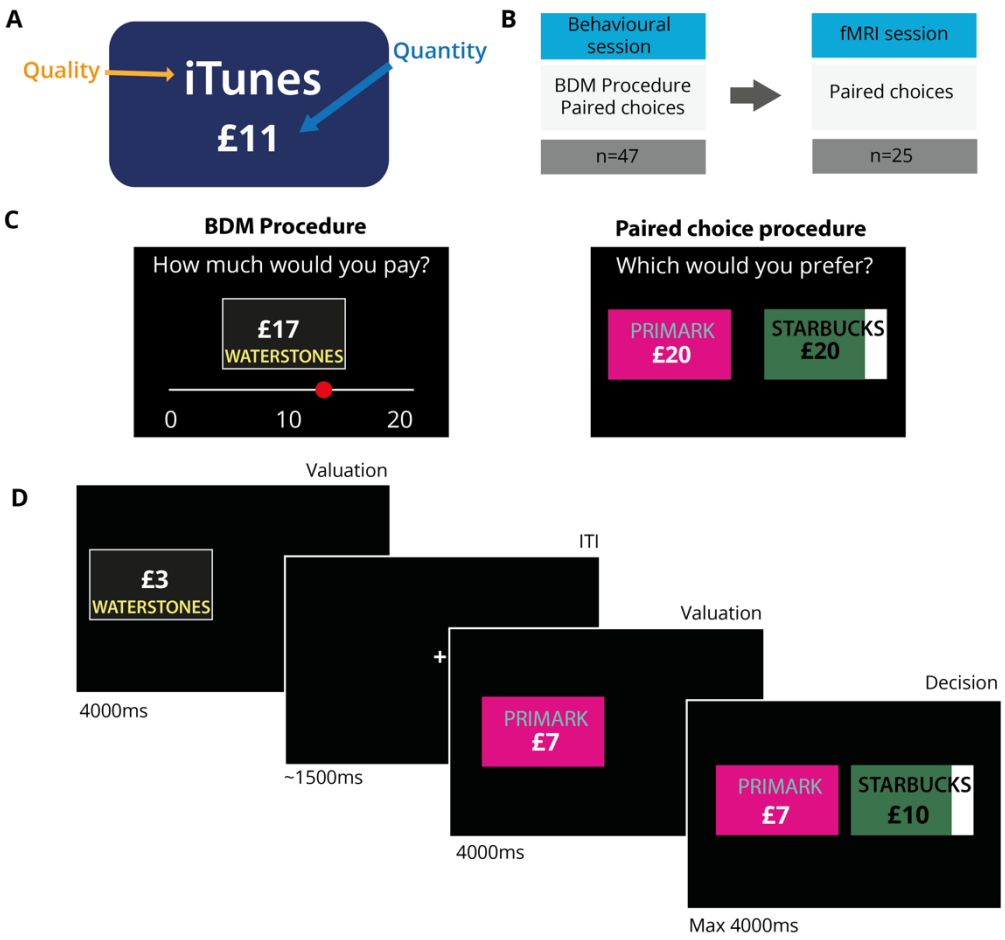


Figure 5.1 | Experimental procedure (A) We used giftcards to manipulate quality and quantity. Cards from different shops had different qualities, depending upon the subjective value of money that can be spent only at that shop. Quantity was varied with the amount of money (number of £) on the card. **(B)** Following an initial behavioural session in which we mapped value functions for different giftcards, a subset of participants were invited to return for an fMRI session. **(C)** Behavioural experiment. The behavioural session comprised two tasks. The first involved an auction procedure (Becker-DeGroot-Marschak procedure, BDM). Participants were offered different cards with varying amounts of money on them, and asked the maximum amount that they would be willing to pay for that card. In the second (paired choice) subjects made choices between pairs of giftcards with equal quantities (£20). **(D)** fMRI experiment. On most trials (6 of 7) participants saw only a single giftcard from one of three different shops, with a randomly varying quantity (amount of money). On decision trials (1 of 7), a second giftcard was displayed 2s after the first, and participants had 4s to make a choice between the two giftcards. ITI's were normally distributed around 1.5s.

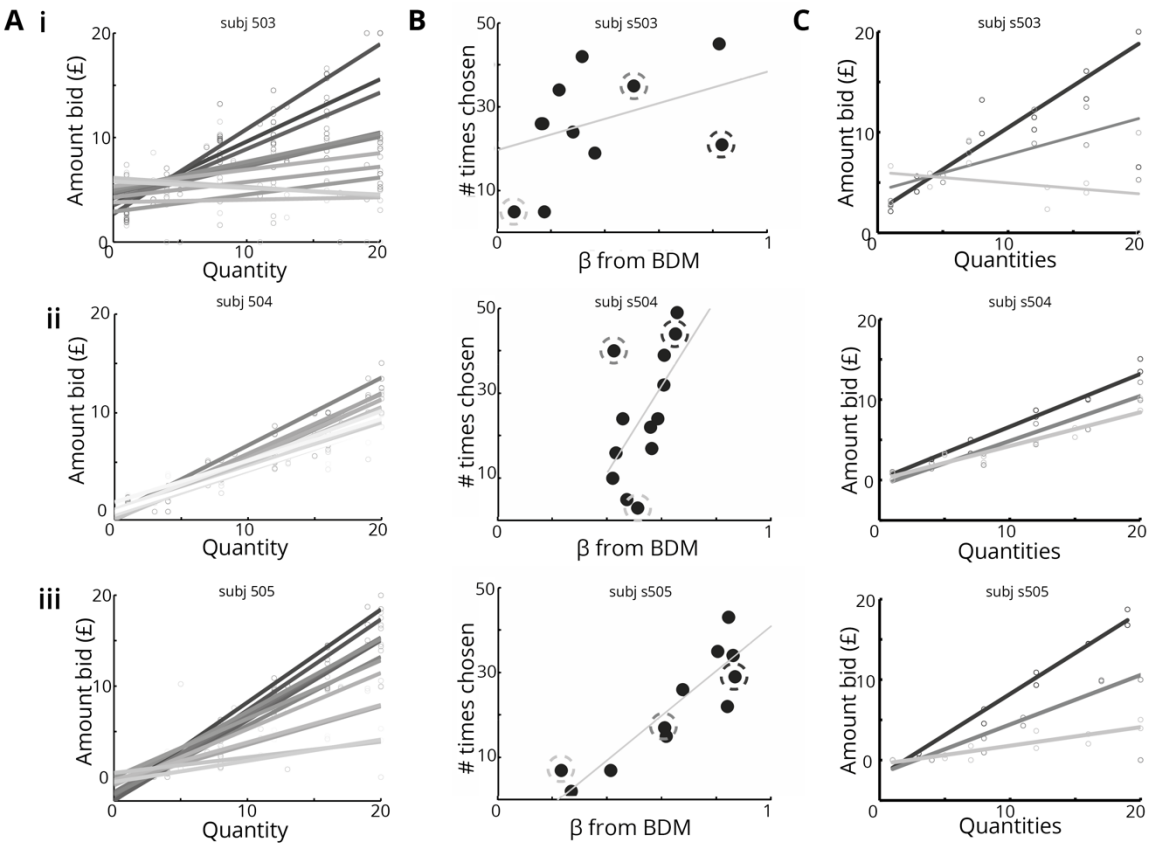


Figure 5.2 | Example participants from behavioural experiment We used data from the behavioural session to determine inclusion in the fMRI study. Our criteria were the consistency and diversity of preferences for different giftcards (assessed using the predictability of BDM ratings), relationships between BDM and paired choice tasks, and correlations between quantity and subjective value in the BDM. **(A)** Firstly, we examined quantity-bid relationships for the 13 different giftcards. The slope of the quantity-bid relationship for each giftcard is a measure of that giftcard's quality, with higher slopes corresponding to more valuable brands. Here, participant i has diverse but noisy preferences, ii is consistent but has similar preferences across giftcards, whilst iii display an acceptable level of consistency whilst maintaining diverse preferences. **(B)** To assess preference stability, we compared the slope of lines estimated from the BDM task with the number of times each giftcard was chosen in the paired-choice task. Participant i shows a weak relationship between choices in each session; ii is consistent but shows little variability; and iii is both consistent and displays diverse preferences. **(C)** For the fMRI experiment, we selected three giftcards for each participant that differed maximally in quality. Here we show BDM plots for selected cards. As before, i is noisy but shows diverse quality preferences, ii has similar preferences over giftcards, and iii has consistent and diverse preferences over giftcards.

5.4.2 fMRI experiment: subjects integrate quality and quantity in choice

For each participant in the fMRI experiment (n=25), we used data from the behavioural session to select 3 giftcards: the giftcard that displayed the steepest relationship between BDM bid and quantity (high quality), the one that displayed the lowest (low quality), and one of intermediate slope (medium quality) (Fig 5.4A). In a pre-scanning paired choice session, we confirmed that preference estimates from the preceding behavioural session were stable, with subjects making choices between the three selected giftcards in a highly predictable manner (Fig 5.4B).

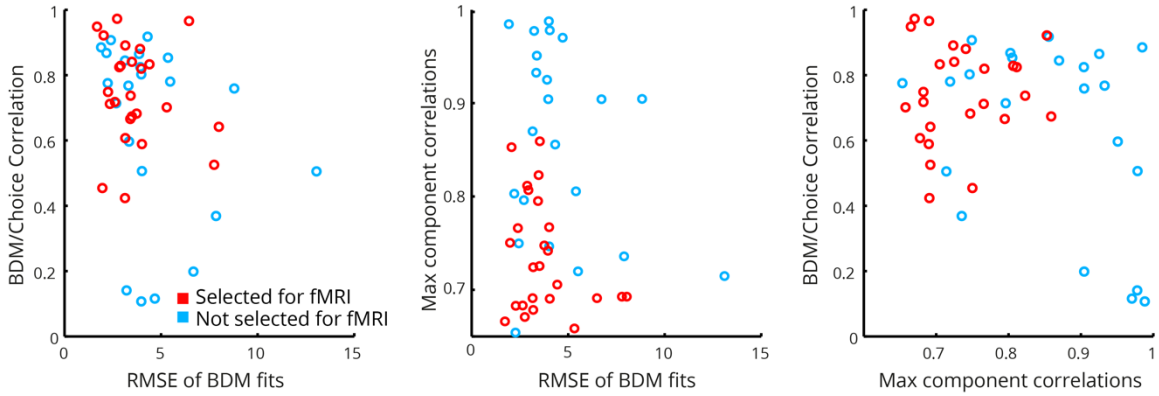


Figure 5.3 | Selected subjects for scanning experiment We selected participants who had high consistency between BDM and paired choice (high BDM/Choice correlation), well-fitting linear models of their BDM ratings (low RMSE of BDM fits), and low correlations between quality, quantity, and integrated value (low component correlations). Overlap between distributions of selected and non-selected subjects reflects the constraints of time and the scanning schedule. Good subjects who were in the same behavioural session as excellent subjects may not have been selected, and mediocre subjects who were in sessions with bad subjects were sometimes selected.

Within the scanner, participants made choices between giftcards of varying quality and quantity on 1/7 of trials (Figure 5.1D), resulting in a total of 48 decisions. We used a logistic regression analysis to quantify the impact of differences between the two options upon choice. We calculated an interaction term as the mean-centred product of quality and quantity. Intuitively, the interaction term captures the fact that an extra £ on the highest quality giftcard is more valuable to the subject than an extra £ on the low quality giftcard. Differences between options in quality, quantity, and the interaction all influenced participants' choices (Quality T_{24}=8.6,

p<0.001; Quantity T_{24}=13.7, p<0.001; Interaction T_{24}=3.8, p<0.001) (Figure 5.4C). Importantly, this implies that participants were performing a multiplication of quality and quantity to estimate integrated value, rather than merely combining them additively.

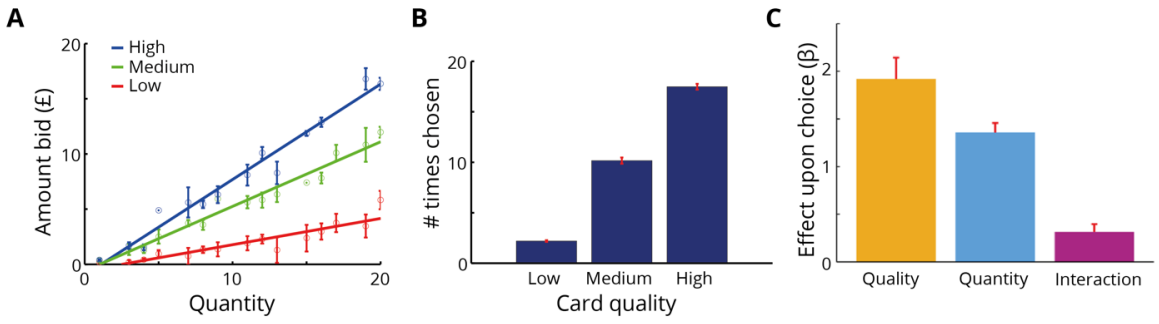


Figure 5.4 | Behavioural results for subjects in scanning experiment (A) Average quantity-bid functions show the difference in quality for three selected giftcards for subjects who completed both the behavioural and fMRI sessions (n=25). **(B)** In a pre-scanning paired-choice session, we confirmed that the ordering of cards by quality was highly consistent between sessions. **(C)** Analysis of choices made in the MRI scanner. During the fMRI experiment, participants made 48 choices between cards of varying quality and quantity (see Fig 5.1D). We used the differences between options to predict choices using logistic regression. The differences between options in quality (T_{24}=8.6, p<0.001) and quantity (T_{24}=13.7,p<0.001) were predictive of choice. Importantly, the interaction between quality and quantity also predicted choice (T_{24}=3.8,p<0.001), consistent with the multiplicative relationship expected from the observed quantity-utility functions (Figure 5.4A). This positive interaction term is consistent with subjects combining quantity and quality into an integrated value.

5.4.3 Brain activity associated with quality, quantity, and their interaction

In the scanner, participants were shown a single giftcard and asked to internally evaluate it (evaluation trials), in the knowledge that they might have to make a fast decision between that option and another (decision trials) (Fig 5.1D). The preponderance of valuation trials (340/388) provided us with an opportunity to examine value computation in isolation, without potentially confounding effects of decision dynamics [49].

In order to isolate the elements of value representation, we used General Linear Models (GLMs) of voxel-wise brain activity to examine the representation of quality, quantity, and their interaction in valuation trials. We split our card presentations by quality (low, medium, high), and associated each onset with a parametric modulator corresponding to the quantity presented on

that trial. This allowed us to look for main effects of card quality, [$Quality_{High}$ - $Quality_{Low}$], card quantity [$Quantity_{LowQuality}$+ $Quantity_{MediumQuality}$+ $Quantity_{HighQuality}$], and the interaction between the two [$Quantity_{HighQuality}$ - $Quantity_{LowQuality}$].

We found three largely non-overlapping networks corresponding to the representation of offer quality, quantity, and their interaction. Higher card quality was associated with activity in the bilateral Inferior Frontal Gyrus (IFG), centred on the pars opecularis (Left: Peak MNI= -54, 12, 30; T_{24}=4.79, $p_{FWE\text{-}corrected}$=0.023; Right: Peak MNI= 51, 9, 27; T_{24}=4.37, $p_{FWE\text{-}corrected}$=0.032) (Fig 5.5A). On the left hand side, this extended into the dorsolateral PFC (Peak MNI= -36, 48, 24; T_{24}=3.16, $p_{FWE\text{-}corrected}$=0.044). Notably, this activation included Broca's area, which is associated with semantic comprehension [94], a critical step in evaluating stimuli such as giftcards. Parameter estimates extracted from group-level functional ROIs (defined at p<0.005) revealed that neither quantity (T_{24}=1.18,p=0.24), nor the interaction (T_{24}=1.53,p=0.14) were associated with increased IFG activity, although direct comparisons did not distinguish quality coding from that of quantity or the interaction (T_{24}=1.88,p=0.073; T_{24}=1.88,p=0.17) .

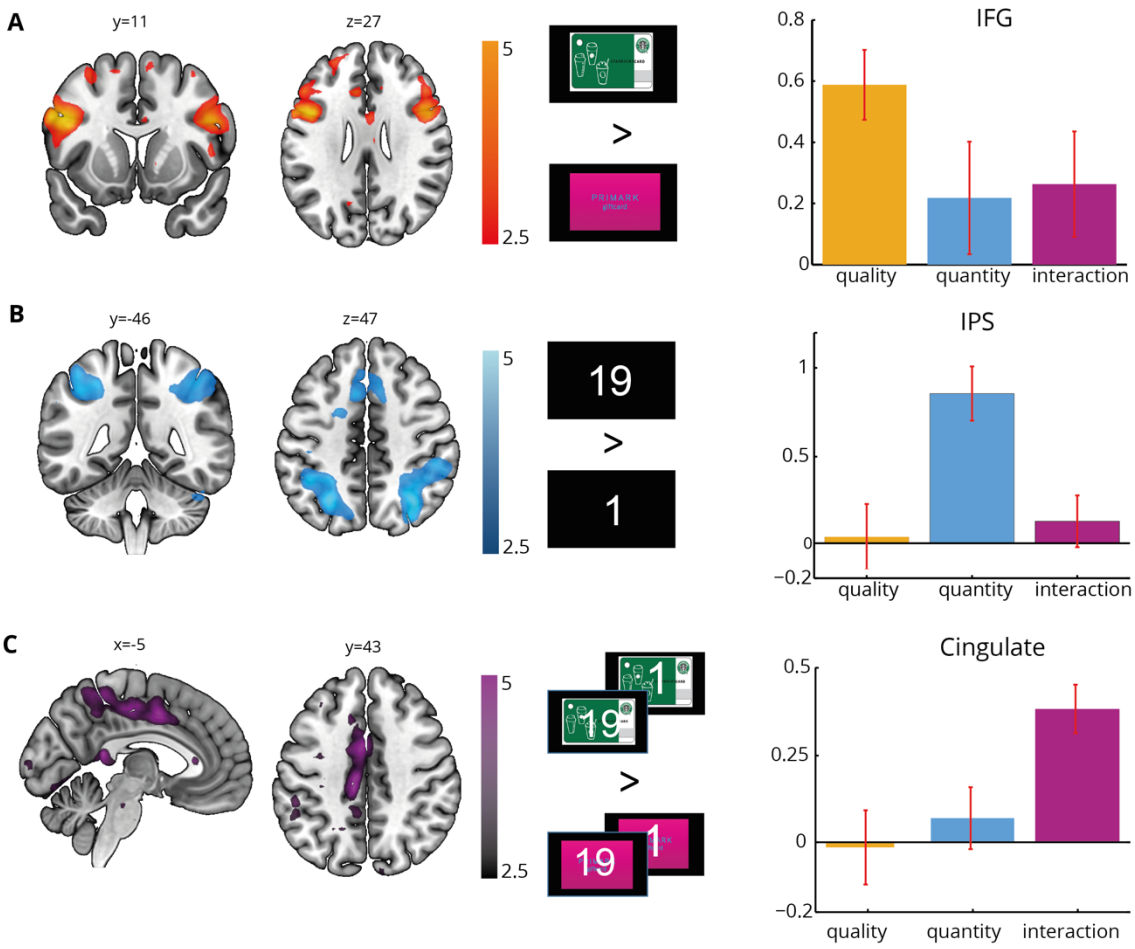


Figure 5.5 | Representation of quality, quantity, and their interaction (A) We observed bilateral coding of offer quality [$Quality_{High}$ – $Quality_{Low}$] bilaterally in the Inferior Frontal Gyrus (IFG). **(B)** Increasing quantity, as tested by [$Quantity_{HighQuality}$ + $Quantity_{MediumQuality}$+$Quantity_{LowQuality}$], was associated with increasing activity bilaterally in the Intra Parietal Sulcus (IPS). **(C)** Activations in the posterior cingulate cortex were consistent with representing the interaction of quality and quantity [$Quantity_{HighQuality}$ – $Quantity_{LowQuality}$]. Errors bars are SEM across subjects, SPMs thresholded at p<0.01 for visualization.

Offer quantity was correlated with activity in bilateral Intraparietal Sulcus (IPS) (Left: Peak MNI: -27,-66,51; T_{24}=4.77, $p_{FWE\text{-}corrected}$<0.001; Right: Peak MNI: 33,-66,51; T_{24}=4.68, $p_{FWE\text{-}corrected}$<0.001) resonating with the well-characterized role of this region in numerical reasoning in humans and non-human primates [67,84,85,95,96]. As with the IFG, activity in the IPS was selective for number, with no obvious coding of quality (T_{24}=0.19,p=0.85) or the interaction (T_{24}=0.84, p=0.41), suggesting

that the IPS is not performing value coding per se, but specifically representing the quantity of available options. Direct comparisons confirmed that quantity correlations were greater than those for quality (T_{24}=3.52, p=0.0017) and for the interaction (T_{24}=3.36,p=0.0025). We also observed quantity-related activity in bilateral visual cortex (Left: Peak MNI: -33,-87,-12; T_{24}=5.54, $p_{FWE\text{-}corrected}$<0.001; Right: Peak MNI: 27,-87,-12; T_{24}=5.49, $p_{FWE\text{-}corrected}$<0.001).

Finally, we asked whether activity in any region of the brain was associated with the interaction between quality and quantity, correlating more steeply with quantity for high than low quality giftcards. This might be considered a signature of value computation, suggesting additional processing above and beyond a simple reflection of option quality and quantity. The most prominent activity associated with this contrast extended along the posterior cingulate (Peak MNI: -12,-15,54; T_{24}=3.57, $p_{FWE\text{-}corrected}$<0.001). As above, activity in this functional ROI was specific for the interaction term, with no evidence of quality (T_{24}=-0.14,p=0.89), or quantity (T_{24}=0.77, p=0.45) correlatinos, implying that despite this region's involvement in value computation, it does not represent an integrated value signal *per se*. Direct comparisons confirmed that interaction exceeded both quality (T_{24}=2.93,p=0.0072) and quantity (T_{24}=2.53,p=0.018) contrasts. Effects of the interaction contrast were also present in bilateral superior temporal lobes (Left: Peak MNI=-63,-45,0; T_{24}=5.55, $p_{FWE\text{-}corrected}$<0.001; Right: Peak MNI=48,-33,3; T_{24}=4.85, $p_{FWE\text{-}corrected}$<0.001).

5.4.4 Computation of integrated value from component parts in the cingulate

Having characterized the neural responses to the individual components of option value (quality, quantity, and their interaction), we next asked whether any regions represented integrated value. Since integrated value is somewhat correlated with quality and quantity (though this correlation is limited by design, see Figure 5.3), merely testing for effects of integrated value presents a problem; regions sensitive to quality or quantity alone might appear to reflect integrated value. To overcome this, we paired a traditional parametric modulation approach (testing for effects of integrated value, quality x quantity), with a conjunction analysis, reasoning that a region truly representing integrated value ought to display sensitivity to all of the component parts: quality, quantity, and their interaction.

Both analyses revealed a striking convergence on the anterior cingulate cortex (ACC) (Figure 6A). Activity in this region covaried with quality, quantity, and their interaction (all $p<0.05_{uncorrected}$), and with a parametric modulator for integrated value (Peak MNI: -12,21,39; $T_{24}=6.04$, $p_{FWE-Corrected}<0.001$). This region is known to contain neurons that multiplex attributes in value-based decision-making [19], and is thought to be crucial to value-learning [53,97].

We further reasoned that if the ACC's value estimates were being used to guide choice, we might see greater activity in decision trials compared to mere valuation trials. This was indeed the case; decision trials were associated with an increase in activity in the same region (Peak MNI=9,15,45; $T_{24}=17.03$, $p_{FWE-Corrected}<0.001$) (Figure 5.6B). Dorsally, this region partially overlaps with activity in dmPFC previously characterized as the final value-comparison step before motor output [98].

Our analyses revealed dissociable representations of quality, in the IFG, and quantity, in the IPS. Although we lack the temporal precision to test whether these segregated representations precede the emergence of integrated value signals in ACC, we can ask whether between-subject variability in component coding is related to between-subject variability in ACC representations. We found that this was indeed the case. As predicted, stronger IFG encoding of quality was associated with stronger coding of quality in the ACC ($r=0.63$, $p<0.001$) (Figure 5.6C), whilst stronger IPS encoding of quantity was associated with stronger quantity coding in the ACC ($r=0.68$, $p<0.001$) (Figure 5.6F). Importantly, the converse correlations did not hold, with parameter estimates for IGF quality unrelated to ACC quantity ($r=0.04$, $p=0.83$) and IPS quantity unrelated to ACC quality ($r=0.14$, $p=0.50$) coding (Figure 5.6D, 5.6E). This specificity suggests that observed correlations reflect meaningful inter-regional relationships rather than correlated variance in signal-to-noise between participants.

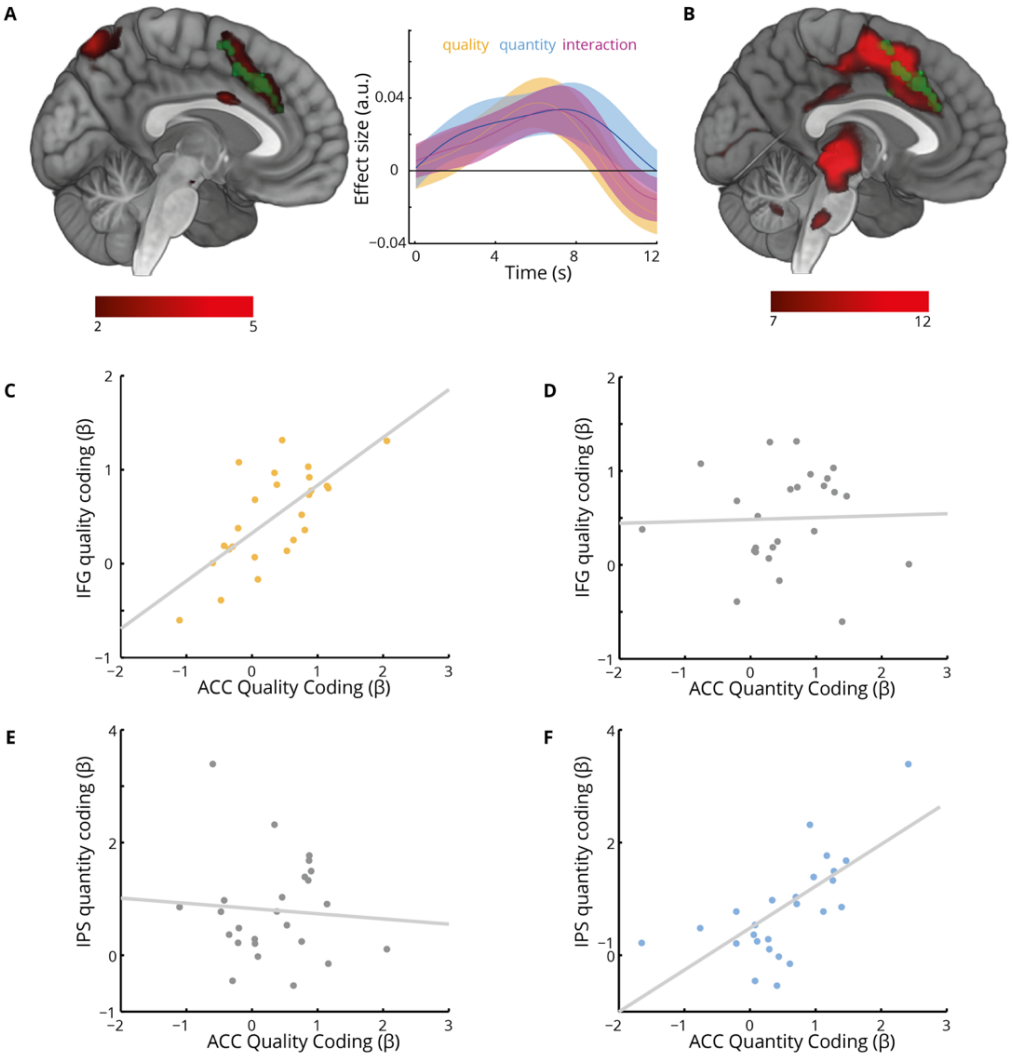


Figure 5.6 | Computation of utility from component parts in the anterior cingulate cortex **(A)** Overlapping effects of quality, quantity, and their interaction in the ACC. A conjunction analysis revealed overlapping representations of each component ($p<0.05_{uncorrected}$) [green] in the ACC, suggesting a nexus for the computation of value. A complementary analysis using an explicit representation of integrated value as a parametric modulator implied the same ($p<0.05_{FWE\text{-}corrected}$) [red]. Timecourse displayed for illustration purposes. **(B)** Decision > Non-Decision trials. The ACC also showed higher activity in trials upon which a decision was made compared to valuation trials [red], overlapping with the conjunction analysis identified in panel a [green]. **(C)** Participants with stronger representations of quality in the IFG showed stronger representations of quality in the ACC (r=0.63, p<0.001). **(D)** Quality sensitivity in IFG was unrelated to quantity coding in ACC (r=0.04, p=0.83). **(E)** Quality sensitivity in IPS was unrelated to quality coding in IPS (r=-0.14, p=0.50). **(F)** Participants with stronger representations of quantity in the IPS showed stronger representations of quantity in the ACC (r=0.68, p<0.001). Each point is one participant.

5.4.5 Strength of neural quantity coding reflects choice predictability

The degree to which subjects' choices were correctly predicted by our logistic regression varied, from 69% to 92%. We reasoned that stronger neural representations of value components might lead to more predictable choices.

Using parameter estimates (β's) extracted from our GLM analyses, we asked whether between-subject variability in β's related to between-subject choice predictability. We found that the strength of neural correlations with quantity, but not quality, were associated with predictability of choice. Mean β's in both the IPS (ρ=0.60, p=0.002) and ACC (ρ=0.42, p=0.039) were positively correlated with choice predictability, suggesting that stronger neural representations of quantity are associated with more reliable choices. Correlations between predictability and quality coding in the IFG (ρ=0.41, p=0.104) and ACC (ρ=0.30, p=0.142) were also positive but did not reach significance, perhaps reflecting the greater range of values in the quantity than the quality domain or the potential for overtraining on giftcard quality.

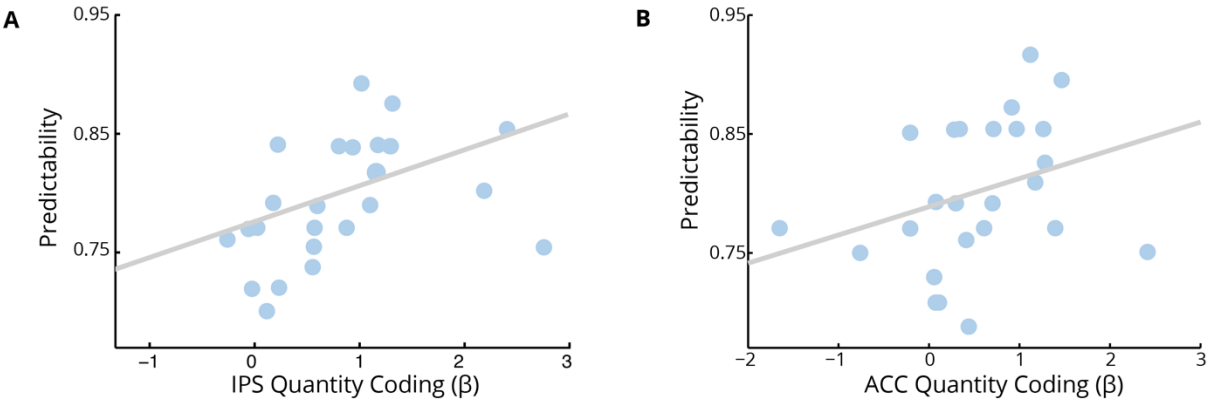


Figure 5.7 | Neural quantity sensitivity relate to choice predictability We found that coefficients for quantity in **(A)** IPS (ρ=0.60, p=0.002) and **(B)** the ACC (ρ=0.42, p=0.039) correlated with the predictability of participants' choices, as assessed by the ability of our logistic regression model to predict choice. Correlations with quality coding in the IFG (ρ=0.41, p=0.104) and ACC (ρ=0.30, p=0.142) were positive but not significant.

5.4.6 Repetition suppression for integrated value in the ACC

The analyses presented thus far provide an insight into the monotonic encoding of value components in the brain; they are sensitive to fluctuations in the BOLD signal which rise and fall in concert with the variables in which we are interested. This approach is common to all but a few recent studies, which have asked how patterns of brain activity might encode value in a more distributed fashion using multi-voxel decoding analyses [17,42,44,99]. A complementary approach, offering insight into neural populations at a sub-voxel resolution [70,71], is to use repetition suppression. Repetition Suppression (RS) describes the phenomenon whereby repeated presentation of stimuli that are similar along some dimension evokes a reduction in activity in brain regions sensitive to that attribute, putatively due to a reduction in activity in neurons that are activated in both trials (Figure 5.8A). This provides a means to assay the neural overlap in the representation of two stimuli, from foods [78], to faces [69], to agents [80]. An influential application of this technique has been to examine numerosity representations in the parietal cortex [84-86]. Using RS designs, activity in the IPS has been found to scale with the absolute difference in number between sequentially presented options, a finding interpreted in the light of Gaussian-tuned number representations in IPS.

Adopting a similar approach, based upon putative populations of tuned neurons spanning the range of integrated values presented in the experiment (Figure 5.8A), we asked whether RS provided additional evidence of non-monotonic value encoding in the ACC. We constructed a GLM in which we modelled the absolute difference in integrated value (ΔIntegratedValue) between subsequent trials, as well as the Integrated Value on each trial. We found evidence for repetition suppression to value in dorsal ACC (Peak MNI: -3,3,51; T_{24}=3.86, $p_{FWE\text{-}Corrected}$=0.003), just posterior to the activity related to monotonic encoding of integrated value (Figure 5.8B). These activations were partially overlapping, such that the integrated-value coding conjunction identified in Figure 5.6 showed effects of both ΔIntegratedValue and Integrated Value (Figure 5.8C) (ΔIntegratedValue: T_{24}=2.48, p=0.020; Integrated Value: T=3.26, p=0.0034). Repetition suppression for integrated value was surprisingly widespread. We also observed repetition suppression for value bilaterally in the lingual gyrus (Left: Peak MNI: -15,-45,-9; T_{24}=6.18, $p_{FWE\text{-}Corrected}$<0.001; Right: Peak MNI: 15,-48,3; T_{24}=4.56, $p_{FWE\text{-}Corrected}$<0.001), the right superior temporal sulcus (Peak MNI: 63,-30,3; T_{24}=5.80, $p_{FWE\text{-}Corrected}$<0.001), and bilaterally in the posterior insula

(Left: Peak MNI: -30,0,-3; T_{24}=5.00, $p_{FWE\text{-}Corrected}$<0.001; Right: Peak MNI: 33,-6,-12; T_{24}=3.60, $p_{FWE\text{-}Corrected}$<0.001)

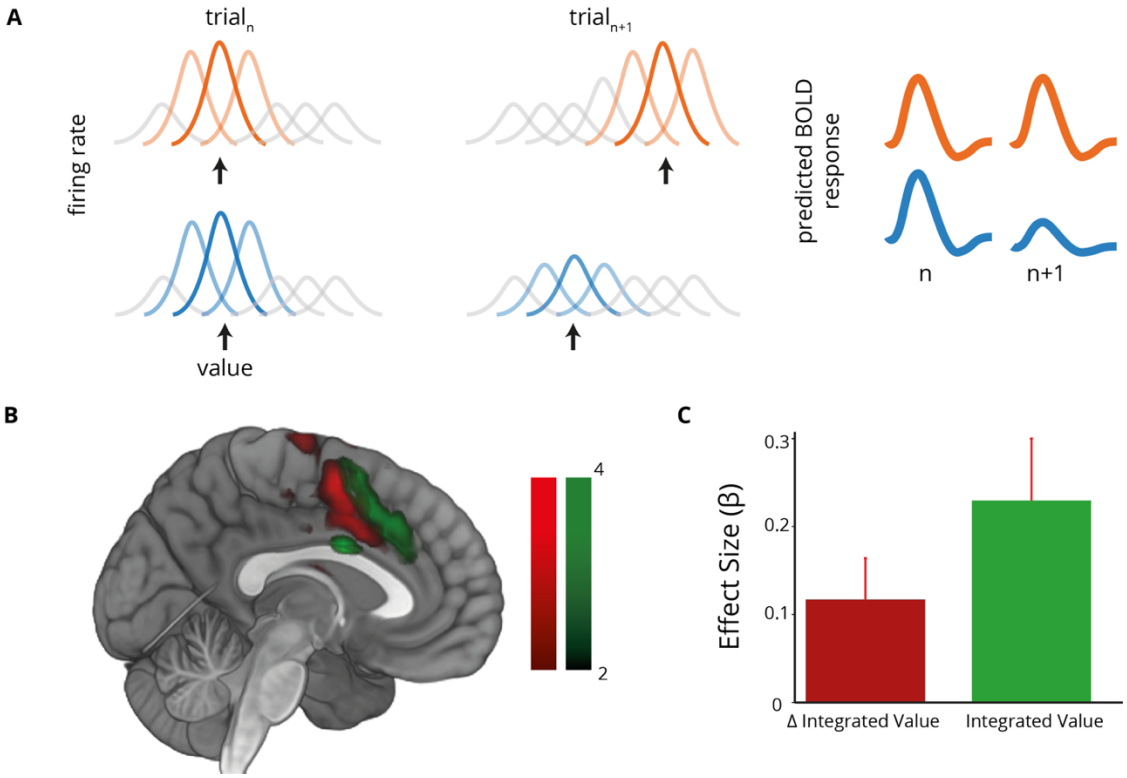


Figure 5.8 | Repetition suppression for value in the anterior cingulate cortex (A) Repetition suppression analysis logic. We hypothesize a population of neurons tuned for value where the different neurons have overlapping tuning curves spanning the range of values presented. Black arrows denote stimulus value for that trial. If consecutive trials activate non-overlapping populations of neurons, evoked responses for each stimulus are similarly high on each trial (top panel in orange). However, repeated presentation of the same stimulus produces repeated activation of the same neurons on consecutive trials, leading to a reduction in the neural response (bottom panel in blue). Summation over all neurons in the population (as in the BOLD signal measured in fMRI), leads to higher activity when consecutive stimuli activate unique subsets of neurons (top panel) than when consecutively activated populations overlap (bottom panel). Predicted BOLD activity is thus proportional to the absolute difference in value between consecutive trials**. (B)** Evidence for multiple forms of value coding in the cingulate. We examined cingulate representations of repetition suppression to integrated value (change in value from trial n-1 to trial n, ΔIntegratedValue) [green], and monotonic encoding of integrated value (a standard parametric modulator approach) [red]. Voxels sensitive to repetition suppression were more posterior, with monotonic encoding stronger in anterior voxels. **(C)** The ACC region identified in the conjunction analysis (Fig 5.6A) also shows repetition suppression to integrated value. We extracted mean parameter estimates for ΔIntegratedValue and for integrated value from the voxels identified in the conjunction analysis. Both were positive on average

(ΔIntegratedValue: T_{24}=2.48,p=0.020; Integrated Value: T=3.26, p=0.0034). Error bars are SEM across participants.

5.4.7 Neural interpretation of repetition suppression effects

Our experiment was designed with reference to numerosity studies, in which, following Piazza et al., [84] repetition suppression effects of the type observed here are interpreted as reflecting the Gaussian tuning curves of the underlying population [100]. It is tempting, therefore, to conclude that our data provide evidence of Gaussian tuning for value in the cingulate, a finding of great interest given theoretical accounts of how tuned populations can multiplex magnitude and uncertainty signals [64,65]. In order to explore the degree to which RS can be used to gain insight into population codes, we constructed a simple firing rate model and asked how aggregate activity in response to a series of 500 cues depended on two factors: tuning curves and adaptation regime.

Our starting point was the model of Piazza et al [84], in which Gaussian neurons undergo divisive adaptation. We modelled a population of 1000 Gaussian neurons of fixed tuning curve width (Figure 9A). In order to avoid end effects, preferred values (i.e. the peak of the Gaussians) were distributed from -15:+45, such that each value in the range 1-30 evoked a similar amount of total activity. On each trial the activity for each neuron was calculated by indexing the tuning curve with the cue value on that trial and adjusting by an adaptation term.

We implemented two forms of adaptation: subtractive, and divisive. Under the subtractive regime, we subtracted a fixed fraction of activity on the previous trial. In the divisive regime, we divided the evoked firing rate by a fixed fraction of activity on the previous trial. In the subtractive regime, therefore, the reduction of activity was independent of the amount of evoked activity on the current trial. Conversely, in the divisive regime, larger firing rates suffered relatively greater adaptation. In order to relate our simulations to BOLD measurements, which sum across many thousands of neurons, we calculated the aggregate activity of the population by taking the mean across all neurons on each trial.

We ran each simulation 100 times, randomizing the tuning curves each time, in order to obtain a distribution of correlation coefficients between aggregate activity and ΔIntegratedValue. Under both adaptation regimes, we confirmed that the aggregate activity in the Gaussian-tuned population grew with the difference in value between consecutive trials (Figure 5.9B & D).

Plotting activity from the run with the median correlation coefficient (Figure 5.9C and 5.9E) demonstrated a clear, albeit non-linear, relationship between ΔIntegratedValue and aggregate activity on each trial. We thus confirmed that our repetition suppression effects were consistent with a Gaussian tuned population irrespective of adaptation regime, confirming and extending previous modelling efforts [84].

We next adapted our model to ask whether we could obtain similar results using a population of cells coding value linearly. We used linear tuning functions of varying slope and intercept, with a mixture of positive and negative slopes for value (Figure 5.9F), consistent with single-neuron recording studies in the ACC [19,23,93,101]. As before, we simulated the response of 100 different populations of cells to the same 500 cues, implemented subtractive or divisive adaptation, and examined the distribution of correlation coefficients between ΔIntegratedValue and aggregate activity.

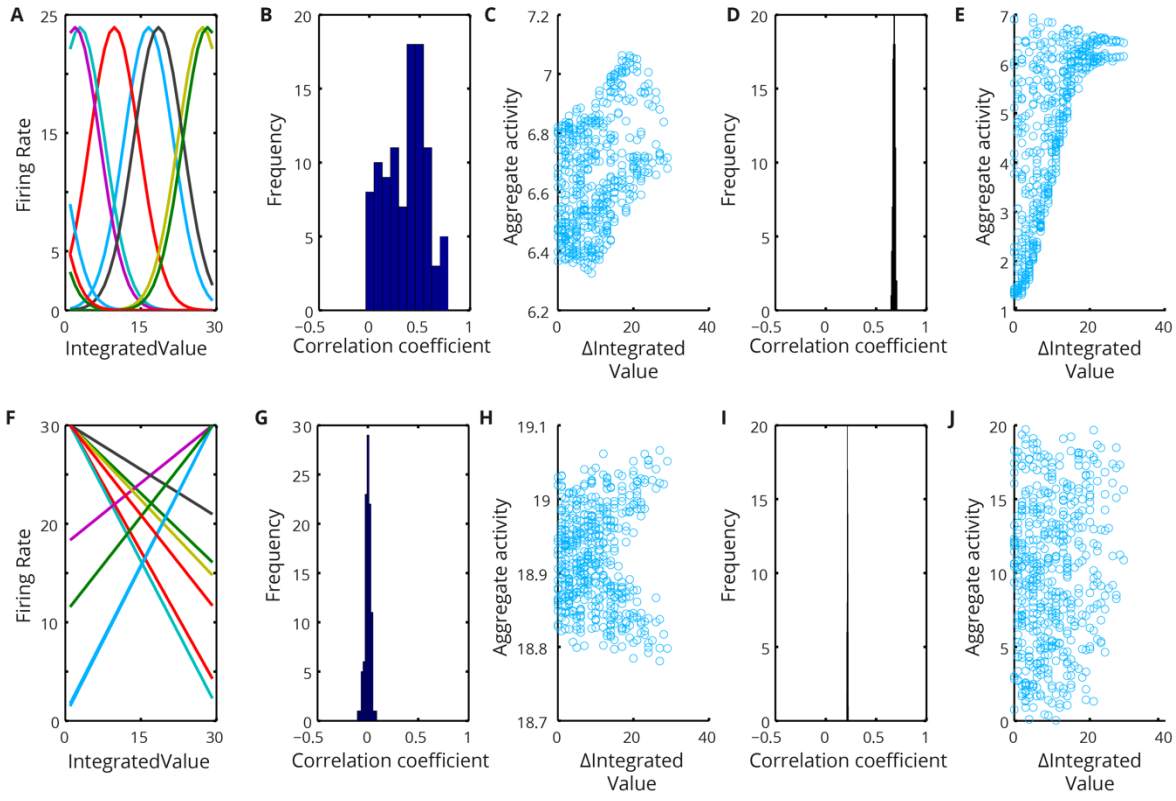


Figure 5.9 | Modelling of repetition suppression: dependence on tuning and adaptation type (A) Tuning curves for value in the 'tuning' model. Tuning curves are Gaussian, with a random mean and fixed width. **(B)** Tuned model with subtractive adaptation: positive

relationship between change in value and activity. **(C)** Example run (median correlation coefficient), showing positive relationship between change in value and activity in tuned subtractive model. Each point is a trial. **(D)** Tuned model with divisive adaptation: positive relationship between ΔIntegratedValue and activity. **(E)** Example run showing positive relationship between change in value and activity in tuned divisive model. Each point is a trial. **(F)** Tuning curve for value in the 'linear' model. Tuning curves are linear, with variable slope and positive intercept. Half the neurons have positive slopes, half negative. **(G)** Linear model with subtractive adaptation: no relationship between ΔIntegratedValue and activity. **(H)** Example run showing no relationship between change in value and activity in linear subtractive model. Each point is a trial. **(I)** Linear model with divisive adaptation: positive relationship between ΔIntegratedValue and activity. **(J)** Example run showing weak positive relationship between ΔIntegratedValue and activity in linear divisive model. Each point is a trial.

We found that our results were strongly influenced by the adaptation regime we employed. Under the subtractive regime, the population displayed no relationship between ΔIntegratedValue and aggregate activity (Figure 5.9G, H). Conversely, if we implemented divisive adaptation, we observed a reliable relationship: trials on which ΔIntegratedValue was higher were associated with higher activity on average (Figure 5.9I, 5.9J). We therefore conclude that repetition suppression effects of the type we observe are not necessarily an unambiguous signature of non-monontonic tuning at the single-neuron level.

In order to understand why the choice of adaptation scheme affected our models' predictions so profoundly, we turned to an explicative toy model consisting of two neurons, one positively tuned (neuron 1), one negatively tuned (neuron 2) (Figure 5.10A). We then contrasted the response to two scenarios: repetition of the same value in two consecutive trials, and a large change in value between trials.

In the subtractive case, the two were equivalent (Figure 5.10B, C). Due to the subtractive nature of the adaptation, it doesn't matter whether the more adapted neuron 1 (which previously experienced a high firing rate) is exposed to a high or low value cue: its firing rate is a fixed quantity lower than the previous trial. Similarly, the minimally adapted neuron 2 fires almost as vigorously as it did in the first trial, regardless of where in its tuning function the stimulus lies. On average, as in the complete model, activity is independent of ΔIntegratedValue (Figure 5.10D).

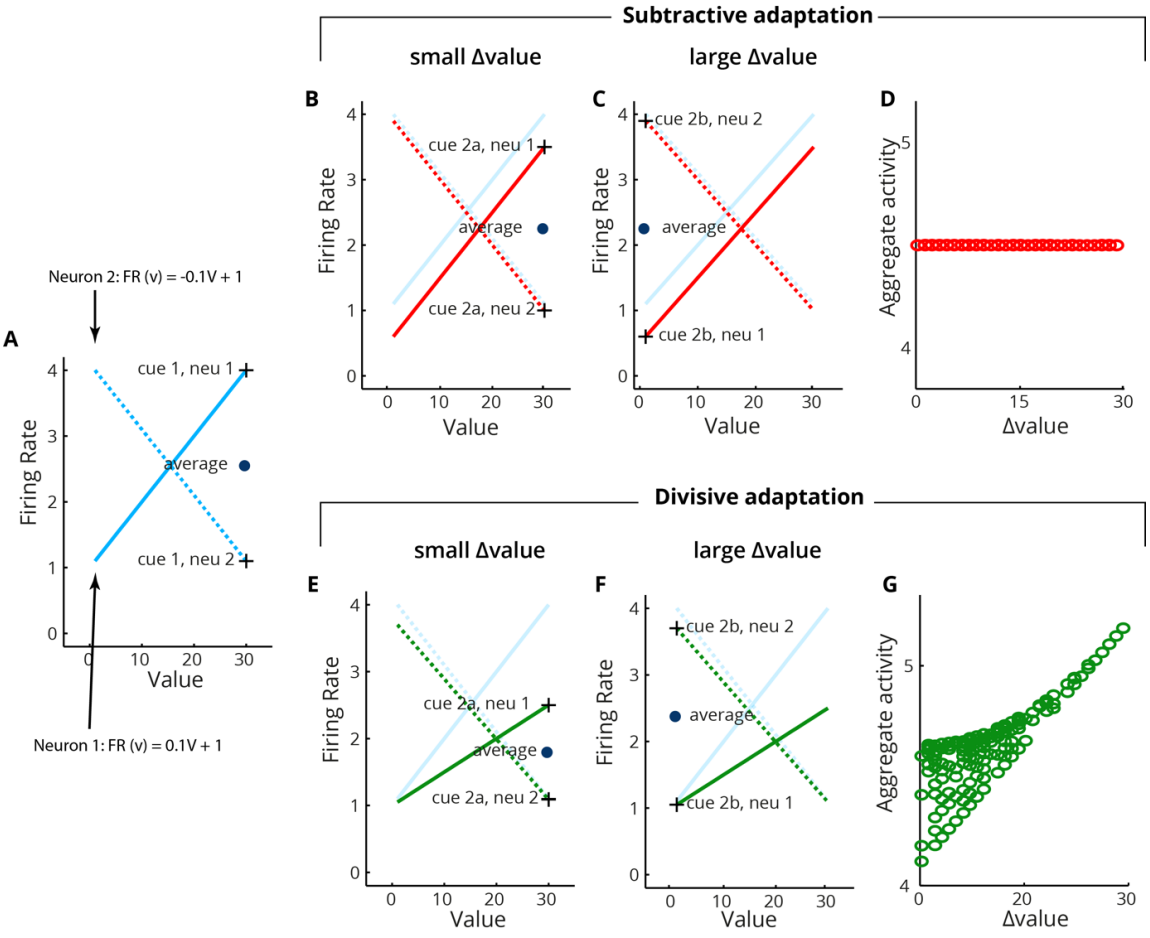


Figure 5.10 | Dissecting divisive vs. subtractive repetition suppression effects For both adaptation schemes we model two neurons with opposite linear tuning to value. In parts A, B,C,E & F, we plot the activation function for each neuron. **(A)** Trial 1. The first cue (value=30) is presented, eliciting high activity in neuron 1 (solid) and low activity in neuron 2 (dashed). **(B)** Trial 2, similar cue (value=30), subtractive adaptation. Evoked activity in neuron 1 is much reduced, whereas neuron 2 is weakly adapted, giving a similar response to trial 1. Aggregate activity is moderate. Tuning curves on previous trial shown in blue for comparison. **(C)** Trial 2, disparate cue (value=1), subtractive adaptation. High activity in the lowly adapted neuron 2 and low in the highly adapted neuron 1. However, since adaptation is subtractive i.e. does not interact with evoked firing rate, aggregate activity matches that in panel b. Tuning curves on previous trial shown in blue for comparison. **(D)** Subtractive adaptation: over 500 trials, aggregate activity is unrelated to the change in value between cues. **(E)** Trial 2, similar cue (value=30), divisive adaptation. Evoked activity in the highly responsive neuron 1 is much reduced, whereas neuron 2 is weakly adapted, giving a similar response to trial 1. Aggregate activity is low. Tuning curves on previous trial shown in blue for comparison. **(F)** Trial 2, disparate cue (value=1), divisive adaptation. High activity in the lowly adapted neuron 2 and low in the highly adapted neuron 1. However, since adaptation is divisive, the impact of the high adaptation of neuron 1 upon the aggregate activity is reduced, leading to higher activity on average. Tuning curves on previous

trial shown in blue for comparison. **(G)** Over 500 trials, aggregate activity scales with change in value between trials.

Conversely, the divisive case produces an *interaction* between current firing rate and previous firing rate. In Figure 5.10E, the cue is repeated, and the heavily adapted neuron 1 produces a response which is substantially attenuated. Neuron 2, which is weakly responsive to the cue, produces a minimally adapted response. Conversely, in 5.10F, when we shift the cue, we increase the firing rate of neuron 2, which displays a much smaller effective adaptation than neuron 1 in 5.10E. This is not outweighed by the adaptation experienced by neuron 1, because the firing rate of neuron 1 is low to that cue. Hence, maximal aggregate activity is seen when consecutive cues occupy opposite ends of the value spectrum, producing a dependence of activity upon ΔIntegratedValue (Figure 5.10G).

Our modelling therefore suggests that our repetition suppression effects are consistent with a Gaussian code, as previously suggested [84]. However, this pattern of activity is not exclusively associated with such a code. A mixed linear code operating with divisive adaptation could also produce the dependence on ΔIntegratedValue that we observe in the ACC (Figure 5.8A).

5.5 Discussion

Value representations are typically studied as monolithic entities. Indeed, considerable effort has been expended in identifying abstract behavioural and neural signatures of scalar value estimates (see sections 1.2 & 1.3). However, recent work suggests that during choice, components of value compete at an attribute-level to guide decisions [102,103], emphasising the importance of decomposing value into its constituent parts. Here we show that in the absence of choice, integrated value correlates appear in the ACC, with component representations in the IFG (quality), and IPS (quantity) (Figures 5.5 & 5.6). A distinct network appears to contribute to the integration of the two, with posterior cingulate and superior temporal lobe activations corresponding to the interaction between quality and quantity (Figure 5.5). We further find that a more posterior region of the ACC displays repetition suppression to integrated value, consistent with (but not conclusive evidence for) multiple forms of value coding in the cingulate (Figure 5.8).

5.5.1 Correlates of quality in the brain

We found that bilateral IFG activity was modulated by the quality of the giftcard presented on each trial (Figure 5.5A). This was unexpected, given the scarcity of reports of IFG involvement in value-based decision making (though see[104-105]). *A priori*, the orbitofrontal cortex (OFC) might represent a more promising candidate for the representation of stimulus quality. However, representations in the OFC appear to be particularly entangled with stimulus identity [38,42,43,78,79], potentially reflecting the central role of the OFC in providing an internal model of the world [106]. This view suggests that the OFC is particularly interested in tracking *relationships* between rewards and their predictors [82,107-111], rather than estimating stimulus quality per se. Furthermore, a recent study found that OFC *exclusively* represented hidden variables related to the current state [112]. The lack of OFC involvement in our task may therefore reflect the static and transparent relationship between stimuli and outcomes in our experiment.

The involvement of the IFG in the representation of stimulus quality is consistent with the semantic nature of the giftcard stimuli we used. IFG is commonly activated in lexical tasks [94], and left-hemisphere lesions to this area famously produce impairments in language production and comprehension. In one of the few studies attempting to parse value into separable components, Lim et al [113] offered participants t-shirts that varied in their aesthetic and semantic properties. They found correlations with aesthetic value in the fusiform gyrus and semantic value in the superior temporal gyrus, whilst vmPFC activity correlated with the value of both attributes. This suggests that the extraction of quality may occur in concert across brain areas specialized for the analysis of different stimulus features, in the same way that feedforward models of visual inputs eventually produce value estimates in deep reinforcement learning networks [114,115]. This suggests that the representation of stimulus quality in IFG may be specific to the use of semantically rich stimuli such as those employed here.

5.5.2 Correlates of quantity in the brain

Conversely, our observation of quantity coding in the IPS (Figure 5.5B) is precisely as predicted from the literature [100]. A wide variety of animals display the ability to make ethologically relevant decisions using number, from lions [116], to crows [117] . Even new-born chicks are capable of tracking the number of an imprinted object that is placed behind a screen [118]. In macaques, such

judgments rely upon a network of frontal and parietal regions containing neurons tuned to different numbers, including the number zero [66-68].

Studies in humans have made use of model-based decoding analyses [96] and repetition suppression designs [84-86] to provide evidence that similar tuning curves for number exist in the human intraparietal sulcus (IPS), although our modelling results suggest that these findings are less conclusive than previously appreciated (see below). Our results imply that the same IPS circuitry subserves number representation in value computation. This is consistent with the recent observation that when number and value are decorrelated, the IPS tracks quantity and not value [119]. This clarifies disparate reports of the role of parietal cortices in value-based decision making, suggesting that when financial stimuli are used [8,120,121], evaluation occurs within a financial framework (such as the BDM) [122,123], or if stimuli merely differ in magnitude [124], parietal responses to quantity may be misconstrued as representing value or its comparison. Conversely, we find that the IPS specifically represents the quantity of an available option, and that the strength of numerical representations in IPS correlates both with choice predictability and ACC quantity coding. This suggests that neurons in the IPS are contributing to the representation of stimulus value in the ACC, and that this representation is subsequently used to guide choice.

5.5.3 Integration of quality and quantity in the ACC

We found that activity in the ACC was consistent with representation of integrated value, displaying not only correlations with quantity and quality but their interaction (Figures 5.5C and 5.6A). This was further supported by a traditional analysis in which integrated value was used as a parametric modulator. Beckmann et al [125] parcellated the cingulate cortex according to connectivity. The region we identify as carrying a integrated value signal corresponds to their region 4, which shows strong connectivity to dorsolateral prefrontal cortex and is commonly implicated in value-based tasks. The region showing repetition suppression effects may be more situated in their region 5, which has a notably higher connectivity to the parietal cortex. This raises the possibility that the repetition suppression we observe in this region is inherited from tuned numerical representations in parietal cortex [67].

The ACC is frequently identified in both human [35,126,127] and animal experiments [19,23,97,101,128] investigating value-based choice. The more dorsal region in which we find signatures of integrated value is associated with tasks in which participants can assign value to particular actions [125]. This is the case in our experiment, since giftcards were displayed either on the left or the right hand side of the screen, such that assessing the value of a particular giftcard was the same as assessing the value of a left/right button press. Dorsal ACC appears to be particularly engaged by foraging type tasks, in which the pertinent comparisons are between options presented sequentially [35-37,126]. Given that our experiment was closely modelled on such tasks, the involvement of the ACC is as predicted.

Our finding that the cingulate integrates information about quality and quantity to form a multiplicative value representation of the current stimulus is also interesting in light of a related literature which has documented the role of the cingulate in the representation of values associated with 'model-based' cognition [129,130]. This somewhat catch-all term describes flexible computation of value associated with a certain stimulus, and is typically contrasted with 'model-free' cognition, in which stimulus or action values are cached and updated only through repeated experience [131]. The multiplication of quantity and quality we observe in the ACC is consistent with the idea that the cingulate provides an internal model which produces estimates of quantities relevant to behavior [30,132,133]. In our case, utility was maximized by combining quality and quantity in a multiplicative manner, and this is what the ACC appears to do, in a manner that reflects the coding of quality and quantity in the frontal and parietal lobes respectively (Figure 5.6 C,F).

We did not observe activity in the ventromedial PFC, the part of the cortex most frequently associated with valuation [134]. This chimes with recent observations suggesting that sequential [135] or speeded [82,135] choices do not engage vmPFC. Whether evaluation alone is an effective engager of vmPFC is also unclear. Early reports suggested that the vmPFC was part of an automatic valuation system [15], but recent work suggests otherwise [16]. vmPFC also fails to represent a value signal in effort-based decisions in humans [139-141]. The few studies that report value-related activity in macaque vmPFC do so in the context of free viewing [136,137] raising the possibility that the vmPFC is particularly engaged when values must be compared via repeated eye movements [22] . The observation that the vmPFC is also crucial for episodic memory and imagination [138-140] ,

and the predominance of saccade-frequency theta oscillations in mPFC [141-143], suggests a more general role for the vmPFC in providing short-term plasticity which allows features – of a scene, an episode, or a choice – to be sewn together over several seconds. This may explain why our task, which required participants to evaluate a single stimulus at a single point in space, did not evoke noticeable vmPFC activity.

5.5.4 Repetition suppression: hints and limitations

We used a repetition suppression design, allowing us to reveal coding schemes hidden to conventional BOLD analyses. We found that parts of the cingulate cortex displayed repetition suppression to integrated value, with activity that scaled with the absolute difference in value between trials (Figure 5.8A). This region was slightly posterior to the peak activity associated with monotonic integrated value, though it did extend into the area identified in the conjunction analysis of quality, quantity, and their interaction (Figure 5.8B). Although our original intention was to follow the precedent from the numerosity-coding literature and provide evidence of non-monotonic tuning for value [84-86,144], subsequent modelling revealed that such repetition suppression effects are not an unambiguous signature of non-monotonic codes, and could arise from mixed-linear coding under divisive adaptation (Figures 5.9 & 5.10).

Our analysis emphasises the importance of careful experimental design for strong interpretations of repetition suppression effects [145]. Recent successes suggest that RS is most effectively used when the format of the code is known *a priori* [81] , to test the relationships *between* representations [78], or to observe changes in representations over time [80]. In the present experiment, careful post-hoc modelling suggested that the signals we observed could be explained as a function either of code or adaptation form, the latter being poorly understood and likely to vary across the brain [145].

Nevertheless, our results are indeed consistent with non-monotonic coding, providing a hint that theoretically-attractive population coding of value [64] might be present in the cortex. We will revisit this in the next chapter, using single-cell recordings to overcome the limitations imposed by non-invasive recordings in humans. To foreshadow our results in that chapter, we find direct evidence for non-linear coding in ACC, confirming the sensitivity, if not the specificity, of our repetition suppression design.

5.5.5 Conclusion

To conclude, we find that a distributed network comprising the intraparietal sulcus, inferior frontal gyrus, and posterior cingulate & superior temporal sulcus contribute to the computation of integrated value in the ACC. The strength of signals in the ACC reflected the degree to which they were represented in brain areas coding for quality (IFG) and quantity (IPS), and stronger brain correlations with quantity were associated with more predictable choices. We further demonstrate that parts of the ACC also show repetition suppression to integrated value, consistent with the idea that tuning for value is non-monotonic in parts of the cortex. Our findings demonstrate how value is assembled from its component parts, and emphasise the potential for repetition suppression as an assay of population encoding scheme.

5.6 References

1. Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience* **9,** 545–556 (2008).
2. Knutson, B., Rick, S., Wimmer, G. E., Prelec, D. & Loewenstein, G. Neural predictors of purchases. *Neuron* **53,** 147–156 (2007).
3. Levy, I., Lazzaro, S. C., Rutledge, R. B. & Glimcher, P. W. Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *Journal of Neuroscience* **31,** 118–125 (2011).
4. Montague, P. R. & Berns, G. S. Neural Economics and the Biological Substrates of Valuation. *Neuron* **36,** 265–284 (2002).
5. Platt, M. L. & Huettel, S. A. Risky business: the neuroeconomics of decision making under uncertainty. *Nature Neuroscience* **11,** 398–403 (2008).
6. O'Doherty, J., Kringelbach, M. & Rolls, E. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature* (2001).
7. FitzGerald, T. H. B., Seymour, B. & Dolan, R. J. The role of human orbitofrontal cortex in value comparison for incommensurable objects. *J. Neurosci.* **29,** 8388–8395 (2009).
8. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience* **10,** 1625–1633 (2007).
9. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441,** 223–226 (2006).
10. Schultz, W., Dayan, P. & al, E. A neural substrate of prediction and reward. *Science* (1997).
11. Kennerley, S., Behrens, T. & al, E. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience* (2011).
12. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Publishing Group* **14,** 933–939 (2011).
13. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27,** 8366–8377 (2007).
14. Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439,** 865–870 (2006).
15. Lebreton, M., Jorge, S., Michel, V., Thirion, B. & Pessiglione, M. An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* **64,** 431–439 (2009).
16. Grueschow, M., Polania, R., Hare, T. A. & Ruff, C. C. Automatic versus Choice-Dependent Value

Representations in the Human Brain. *Neuron* **85,** 874–885 (2015).
17. Vaidya, A. R. & Fellows, L. K. Ventromedial Frontal Cortex Is Critical for Guiding Attention to Reward-Predictive Visual Features in Humans. *J. Neurosci.* **35,** 12813–12823 (2015).
18. Dayan, P. Instrumental vigour in punishment and reward - Dayan - 2012 - European Journal of Neuroscience - Wiley Online Library. *European Journal of Neuroscience* (2012).
19. Loh, E. *et al.* Context-specific activation of hippocampus and SN/VTA by reward is related to enhanced long-term memory for embedded objects. *Neurobiology of learning and memory* **134 Pt A,** 65–77 (2016).
20. Padoa-Schioppa, C. Neurobiology of economic choice: a good-based model. *Annual Review of Neuroscience* **34,** 333–359 (2011).
21. Polania, R., Krajbich, I., Grueschow, M. & Ruff, C. C. Neural Oscillations and Synchronization Differentially Support Evidence Accumulation in Perceptual and Value-Based Decision Making. *Neuron* **82,** 709–720 (2014).
22. Krajbich, I., Armel, C. & Rangel, A. Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience* **13,** 1292–1298 (2010).
23. Hunt, L. *et al.* Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience* **15,** 470–6– S1–3 (2012).
24. Rustichini, A. & Padoa-Schioppa, C. A neuro-computational model of economic decisions. *J. Neurophysiol.* **114,** 1382–1398 (2015).
25. Chau, B. K. H., Kolling, N., hunt, L. T., Walton, M. E. & Rushworth, M. F. S. A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nature Neuroscience* **17,** 463–470 (2014).
26. Wang, X. J. Decision Making in Recurrent Neuronal Circuits. *Neuron* (2008).
27. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (1998).
28. Kahnt, T., Heinzle, J., Park, S. Q. & Haynes, J. D. Decoding the Formation of Reward Predictions across Learning. *J. Neurosci.* **31,** 14624–14630 (2011).
29. Rutledge, R. B., Dean, M., Caplin, A. & Glimcher, P. W. Testing the reward prediction error hypothesis with an axiomatic model. *J. Neurosci.* **30,** 13525–13536 (2010).
30. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442,** 1042–1045 (2006).
31. Padoa-Schioppa, C. & Schoenbaum, G. Dialogue on economic choice, learning theory, and neuronal representations. *Current Opinion in Behavioral Sciences* **5,** 16–23 (2015).
32. O'Doherty, J. P. The problem with value. *Neurosci Biobehav Rev* **43,** 259–268 (2014).
33. Kolling, N., Wittmann, M. & Rushworth, M. F. S. Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* **81,** 1190–1202 (2014).
34. Rutledge, R. B. *et al.* Risk Taking for Potential Reward Decreases across the Lifespan. *Curr. Biol.* **26,** 1634–1639 (2016).
35. Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience* **33,** 269–298 (2010).
36. Sugrue, L. P. Matching Behavior and the Representation of Value in the Parietal Cortex. *Science* **304,** 1782–1787 (2004).
37. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural mechanisms of foraging. *Science* **336,** 95–98 (2012).
38. Kolling, N., Behrens, T., Wittmann, M. K. & Rushworth, M. Multiple signals in anterior cingulate cortex. *Current Opinion in Neurobiology* **37,** 36–43 (2016).
39. Padoa-Schioppa, C. & Assad, J. A. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience* **11,** 95–102 (2008).
40. Xie, J. & Padoa-Schioppa, C. Neuronal remapping and circuit persistence in economic decisions. *Nature Publishing Group* **19,** 855–861 (2016).
41. De Martino, B., Fleming, S. M., Garrett, N. & Dolan, R. J. Confidence in value-based choice. *Nature Publishing Group* **16,** 105–110 (2013).

42. Howard, J. D., Gottfried, J. A., Tobler, P. N. & Kahnt, T. Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proceedings of the National Academy of Sciences* **112,** 5195–5200 (2015).
43. McNamee, D., Rangel, A. & O'Doherty, J. P. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Publishing Group* **16,** 479–485 (2013).
44. Kahnt, T., Heinzle, J., Park, S. Q. & Haynes, J. D. The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences* **107,** 6010–6015 (2010).
45. Sokol-Hessner, P. *et al.* Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences* **106,** 5035–5040 (2009).
46. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences* **111,** 12252–12257 (2014).
47. Wright, N. D., Symmonds, M. & Dolan, R. J. Distinct encoding of risk and value in economic choice between multiple risky options. *Neuroimage* **81,** 431–440 (2013).
48. Becker, G. M., DeGroot, M. H. & Marschak, J. Measuring utility by a single-response sequential method. *Behav Sci* **9,** 226–232 (1964).
49. Hunt, L. T., Behrens, T. E. J., Hosokawa, T., Wallis, J. D. & Kennerley, S. W. Capturing the temporal evolution of choice across prefrontal cortex. *eLife Sciences* **4**, (2015).
50. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503,** 78–84 (2013).
51. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497,** 585–590 (2013).
52. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology* **37,** 66–74 (2016).
53. Rushworth, M. F. S. & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience* **11,** 389–397 (2008).
54. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature Neuroscience* **10,** 1214–1221 (2007).
55. Meyniel, F., Sigman, M. & Mainen, Z. F. Perspective. *Neuron* **88,** 78–92 (2015).
56. Monosov, I. E. & Hikosaka, O. Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nature Publishing Group* **16,** 756–762 (2013).
57. Kepecs, A., Uchida, N., Zariwala, H. & al, E. Neural correlates, computation and behavioural impact of decision confidence. *Nature* (2008).
58. Schultz, W., O'Neill, M., Tobler, P. N. & Kobayashi, S. Neuronal signals for reward risk in frontal cortex. *Annals of the New York Academy of Sciences* **1239,** 109–117 (2011).
59. Hsu, M., Bhatt, M., Adolphs, R., Tranel, D. & Camerer, C. F. Neural systems responding to degrees of uncertainty in human decision-making. *Science* **310,** 1680–1683 (2005).
60. O'Neill, M. & Schultz, W. Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron* **68,** 789–800 (2010).
61. Bach, D. R. & Dolan, R. J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience* **13,** 572–586 (2012).
62. Preuschoff, K., Bossaerts, P. & Quartz, S. R. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* **51,** 381–390 (2006).
63. Critchley, H. D., Mathias, C. J. & Dolan, R. J. Neural Activity in the Human Brain Relating to Uncertainty and Arousal during Anticipation. *Neuron* **29,** 537–545 (2001).
64. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nature Neuroscience* **9,** 1432–1438 (2006).
65. Beck, J. M. *et al.* Probabilistic population codes for Bayesian decision making. *Neuron* **60,** 1142–1152 (2008).
66. Nieder, A., Freedman, D. J. & Miller, E. K. Representation of the quantity of visual items in the primate prefrontal cortex. *Science* **297,** 1708–1711 (2002).

67. Nieder, A. & Miller, E. K. A parieto-frontal network for visual numerical information in the monkey. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 7457–7462 (2004).
68. Ramirez-Cardenas, A., Moskaleva, M. & Nieder, A. Neuronal Representation of Numerosity Zero in the Primate Parieto-Frontal Number Network. *Current Biology* **26,** 1285–1294 (2016).
69. Loffler, G., Yourganov, G., Wilkinson, F. & Wilson, H. R. fMRI evidence for the neural representation of faces. *Nature Neuroscience* **8,** 1386–1390 (2005).
70. Grill-Spector, K., Henson, R. & Martin, A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci. (Regul. Ed.)* **10,** 14–23 (2006).
71. Krekelberg, B., Boynton, G. M. & van Wezel, R. J. A. Adaptation: from single cells to BOLD signals. *Trends in neurosciences* **29,** 250–256 (2006).
72. Sawamura, H. Using Functional Magnetic Resonance Imaging to Assess Adaptation and Size Invariance of Shape Processing by Humans and Monkeys. *J. Neurosci.* **25,** 4294–4306 (2005).
73. Sawamura, H., Orban, G. A. & Vogels, R. Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. *Neuron* **49,** 307–318 (2006).
74. Benucci, A., Saleem, A. B. & Carandini, M. Adaptation maintains population homeostasis in primary visual cortex. *Nature Publishing Group* **16,** 724–729 (2013).
75. Kar, K. & Krekelberg, B. Testing the assumptions underlying fMRI adaptation using intracortical recordings in area MT. *Cortex* **80,** 21–34 (2016).
76. Pedreira, C. *et al.* Responses of Human Medial Temporal Lobe Neurons Are Modulated by Stimulus Repetition. *J. Neurophysiol.* **103,** 97–107 (2010).
77. Li, L., Miller, E. K. & Desimone, R. The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* **69,** 1918–1929 (1993).
78. Barron, H. C., Dolan, R. J. & Behrens, T. E. J. Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience* **16,** 1492–1498 (2013).
79. Klein-Flügge, M. C., Barron, H. C., Brodersen, K. H., Dolan, R. J. & Behrens, T. E. J. Segregated Encoding of Reward-Identity and Stimulus-Reward Associations in Human Orbitofrontal Cortex. *J. Neurosci.* **33,** 3202–3211 (2013).
80. Garvert, M. M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T. E. J. & Dolan, R. J. Learning-Induced Plasticity in Medial Prefrontal Cortex Predicts Preference Malleability. *Neuron* **85,** 418–428 (2015).
81. Doeller, C. F., Barry, C. & Burgess, N. Evidence for grid cells in a human memory network. *Nature* **463,** 657–661 (2010).
82. Boorman, E. D., Rajendran, V. G., O'Reilly, J. X. & Behrens, T. E. Two Anatomically and Computationally Distinct Learning Signals Predict Changes to Stimulus-Outcome Associations in Hippocampus. *Neuron* **89,** 1343–1354 (2016).
83. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352,** 1464–1468 (2016).
84. Piazza, M., Izard, V., Pinel, P., Le Bihan, D. & Dehaene, S. Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* **44,** 547–555 (2004).
85. Piazza, M., Pinel, P., Le Bihan, D. & Dehaene, S. A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* **53,** 293–305 (2007).
86. Jacob, S. N. & Nieder, A. Notation-Independent Representation of Fractions in the Human Parietal Cortex. *J. Neurosci.* **29,** 4652–4657 (2009).
87. Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M. & Egner, T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience* **11,** 1004–1006 (2008).
88. Weiskopf, N., Hutton, C., Josephs, O. & Deichmann, R. Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. *Neuroimage* **33,** 493–504 (2006).
89. Deichmann, R., Schwarzbauer, C. & Turner, R. Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. *Neuroimage* **21,** 757–767 (2004).
90. C Hutton et al. The impact of physiological noise correction on fMRI at 7 T. *Neuroimage* **57,** 101

(2011).
91. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38,** 95–113 (2007).
92. Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C. & Worsley, K. J. Multisubject fMRI studies and conjunction analyses. *Neuroimage* **10,** 385–396 (1999).
93. Amemori, K.-I. & Graybiel, A. M. Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nature Neuroscience* **15,** 776–785 (2012).
94. Price, C. J. A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62,** 816–847 (2012).
95. Pinel, P., Piazza, M., Le Bihan, D. & Dehaene, S. Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* **41,** 983–993 (2004).
96. Harvey, B. M., Klein, B. P., Petridou, N. & Dumoulin, S. O. Topographic representation of numerosity in the human parietal cortex. *Science* **341,** 1123–1126 (2013).
97. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Fictive reward signals in the anterior cingulate cortex. *Science* **324,** 948–950 (2009).
98. Hare, T. A., Schultz, W., Camerer, C. F., O'Doherty, J. P. & Rangel, A. Transformation of stimulus value signals into motor commands during simple choice. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 18120–18125 (2011).
99. Kahnt, T., Park, S. Q., Haynes, J. D. & Tobler, P. N. Disentangling neural representations of value and salience in the human brain. *Proceedings of the National Academy of Sciences* (2014). doi:10.1073/pnas.1320189111
100. Nieder, A. The neuronal code for number. *Nature Reviews Neuroscience* (2016).
101. Cai, X. & Padoa-Schioppa, C. Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J. Neurosci.* **32,** 3791–3808 (2012).
102. Zysset, S. *et al.* The neural implementation of multi-attribute decision making: a parametric fMRI study with human subjects. *Neuroimage* **31,** 1380–1388 (2006).
103. Hunt, L. T., Dolan, R. J. & Behrens, T. E. J. Hierarchical competitions subserving multi-attribute choice. *Nature Neuroscience* **17** 1–14 (2014). doi:10.1038/nn.3836
104. Rogers, R. D. *et al.* Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex. *J. Neurosci.* **19,** 9029–9038 (1999).
105. Liljeholm, M., Tricomi, E., O'Doherty, J. P. & Balleine, B. W. Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. *J. Neurosci.* **31,** 2474–2480 (2011).
106. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron* **81,** 267–279 (2014).
107. Stalnaker, T. A., Cooch, N. K. & Schoenbaum, G. What the orbitofrontal cortex does not do. *Nature Publishing Group* **18,** 620–627 (2015).
108. Lopatina, N. *et al.* Lateral orbitofrontal neurons acquire responses to upshifted, downshifted, or blocked cues during unblocking. *eLife Sciences* **4,** (2015).
109. Stalnaker, T. A. *et al.* Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nat Comms* **5,** 3926 (2014).
110. Takahashi, Y. K. *et al.* Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron* **80,** 507–518 (2013).
111. Lucantonio, F. *et al.* Neural Estimates of Imagined Outcomes in Basolateral Amygdala Depend on Orbitofrontal Cortex. *J. Neurosci.* **35,** 16521–16530 (2015).
112. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* **91,** 1402–1412 (2016).
113. Lim, S.-L., O'Doherty, J. P. & Rangel, A. Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *J. Neurosci.* **33,** 8729–8741 (2013).
114. Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518,** 529–533

(2015).
115. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529,** 484–489 (2016).
116. McComb, K., Packer, C. & Pusey, A. Roaring and numerical assessment in contests between groups of female lions, Panthera leo. *Animal Behaviour* **47,** 379–387 (1994).
117. Nor Amira Abdul Rahman. The Numerical Competency of Two Bird Species (Corvus splendens and Acridotheres tristis). *Tropical Life Sciences Research* **25,** 95 (2014).
118. Rugani, R., Fontanari, L., Simoni, E., Regolin, L. & Vallortigara, G. Arithmetic in newborn chicks. *Proceedings of the Royal Society B: Biological Sciences* rspb.2009.0044 (2009). doi:10.1098/rspb.2009.0044
119. Kanayet, F. J., Opfer, J. E. & Cunningham, W. A. The Value of Numbers in Economic Rewards. *Psychol Sci* **25,** 1534–1545 (2014).
120. Clithero, J. A., Carter, R. & Huettel, S. A. Local pattern classification differentiates processes of economic valuation. *Neuroimage* (2009).
121. Ballard, K. & Knutson, B. Dissociable neural representations of future reward magnitude and delay during temporal discounting. *Neuroimage* (2009).
122. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27,** 9984–9988 (2007).
123. Medic, N. *et al.* Dopamine modulates the neural representation of subjective value of food in hungry subjects. *J. Neurosci.* **34,** 16856–16864 (2014).
124. Louie, K., Grattan, L. E. & Glimcher, P. W. Reward Value-Based Gain Control: Divisive Normalization in Parietal Cortex. *J. Neurosci.* **31,** 10627–10639 (2011).
125. Beckmann, M. & Johansen-Berg, H. *Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization*. (The Journal of …, 2009).
126. Boorman, E. D., Rushworth, M. F. & Behrens, T. E. Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. *J. Neurosci.* **33,** 2242–2253 (2013).
127. Bush, G., Vogt, B. A., Holmes, J. & Dale, A. M. Dorsal anterior cingulate cortex: a role in reward-based decision making. in (2002).
128. Hayden, B. Y. & Platt, M. L. Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* **30,** 3339–3346 (2010).
129. Wunderlich, K., Dayan, P. & Dolan, R. J. Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience* **15,** 786–791 (2012).
130. Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D. & Daw, N. D. Model-based choices involve prospective neural activity. *Nature Publishing Group* **18,** 767–772 (2015).
131. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80,** 312–325 (2013).
132. Economides, M., Guitart-Masip, M., Kurth-Nelson, Z. & Dolan, R. J. Anterior cingulate cortex instigates adaptive switches in choice by integrating immediate and delayed components of value in ventromedial prefrontal cortex. *J. Neurosci.* **34,** 3340–3349 (2014).
133. O'Reilly, J. X. *et al.* Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences* **110,** E3660–9 (2013).
134. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76,** 412–427 (2013).
135. Jocham, G. *et al.* Dissociable contributions of ventromedial prefrontal and posterior parietal cortex to value-guided choice. *Neuroimage* **100,** 498–506 (2014).
136. Abitbol, R. *et al.* Neural mechanisms underlying contextual dependency of subjective values: converging evidence from monkeys and humans. *J. Neurosci.* **35,** 2308–2320 (2015).
137. Strait, C. E., Blanchard, T. C. & Hayden, B. Y. Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron* **82,** 1357–1366 (2014).
138. Benoit, R. G., Szpunar, K. K. & Schacter, D. L. Ventromedial prefrontal cortex supports affective

future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences* **111,** 16550–16555 (2014).
139. Hassabis, D. & Maguire, E. A. The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364,** 1263–1271 (2009).
140. Bertossi, E., Tesini, C., Cappelli, A. & Ciaramelli, E. Ventromedial prefrontal damage causes a pervasive impairment of episodic memory and future thinking. *Neuropsychologia* (2016).
141. Guitart-Masip, M. *et al.* Synchronization of Medial Temporal Lobe and Prefrontal Rhythms in Human Decision Making. *J. Neurosci.* **33,** 442–451 (2013).
142. Adhikari, A., Topiwala, M. A. & Gordon, J. A. Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety. *Neuron* **65,** 257–269 (2010).
143. Paz, R., Bauer, E. P. & Paré, D. Theta synchronizes the activity of medial prefrontal neurons during learning. *Learn. Mem.* **15,** 524–531 (2008).
144. Ansari, D. & Dhital, B. Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: an event-related functional magnetic resonance imaging study. *J Cogn Neurosci* (2006).
145. Barron, H. C., Garvert, M. M. & Behrens, T. E. J. Repetition suppression: a means to index neural representations using BOLD? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **371,** (2016).

Chapter 6: Diversity of value-tuning in primate prefrontal cortex

The data reported in this chapter were collected by Nish Malalasekera, Laurence Hunt, and Steve Kennerley, to whom I am very grateful.

6.1 Abstract

In the previous chapter we showed that fMRI signal in the Anterior Cingulate Cortex (ACC) during a foraging-type task showed repetition suppression consonant with non-linear value-tuning. This suggested that the ACC might be a candidate for probabilistic population codes for value. However, neural modelling suggested that this signal was not an unambiguous signature of non-linear tuning. In this chapter we ask a similar question using a more direct readout of neural selectivity. Taking extracellular electrophysiological recordings from macaque monkeys, we look for evidence of non-linear coding in single cells in four areas of the prefrontal cortex; the ACC and the orbitofrontal cortex (OFC), lateral prefrontal cortex (LPFC), and ventromedial prefrontal cortices (vmPFC). We find evidence of non-linear tuning in the ACC and the OFC, providing the first evidence that these regions may employ population codes to represent the value of stimuli.

6.2 Introduction

As discussed in section 1.4.6, probabilistic codes provide a powerful way to represent entire probability distributions over some variable, rather than a point estimate[1]. The PPC model as proposed by Ma [2] further allows evidence integration in a Bayesian fashion by summation of codes providing optimal learning with a simple linear operation. This emerges from the representation of precision in the amplitude of population responses, such that more precise (less uncertain) representations are naturally favoured (see Figure 1.7). This is in stark contrast to the 'summary statistic' representations typically found in the neuroeconomic literature, in which single neurons are often found to represent the parameters of a reward distribution, such as its mean and variance (see section 5.2.2). Apart from the difficulty of binding these pieces of information for the purpose of computation, this format places clear constraints upon the probability distribution that neurons can represent. The form of the distribution must be pre-specified, in order for representation by summary statistic to be useful. By contrast, PPCs can be used to represent complex and arbitrary probability distributions [3], such as those encountered in the real world (see section 1.4.6).

In this chapter we seek further evidence of non-linear value representations that might underlie PPCs for value. Building upon the observation in the previous chapter that repetition suppression in the Anterior Cingulate Cortex (ACC) is consistent with non-monotonic tuning for value, we start by reviewing the prevalence of non-linear tuning in the brain, before discussing why ACC and OFC are particularly plausible sites for population coding of value. We then present data recorded from macaque prefrontal cortex which provides direct evidence for non-linear coding of value.

Population codes typically decompose the range of stimuli to be represented into the firing rates of multiple cells, such that any one cell is responsive to a subset of possible stimulus values. For instance, cells tuned to orientation in visual cortex respond to a restricted range of line orientations, displaying narrow Gaussian tuning curves [4] (Figure 6.1A)*. A given stimulus thus

* Figure 6.1A and 6.1B portray idealized tuning curves from a neural network model of object localization, but provide a useful visualization of tuning curves observed elsewhere, such as in V1

evokes activity in the subset of cells with preferred tuning near the presented orientation (Figure 6.1B). Such non-linear, and typically Gaussian, tuning is widely documented in other brain regions, such as higher [5] visual areas, auditory cortices [1,6], and motoric frontal areas [2,7]. Tuning for number is observed in the intra-parietal sulcus and PFC [3,8,9] (see Figure 6.6), which suggests that tuned codes for highly-processed information can be found in areas also involved in choice [4,10]. Linear coding certainly appears to be the exception, rather than the norm, in the brain. Given a recent flurry of reports that PFC coding is more diverse [5,11,12], distributed [12], and labile [13] than previously appreciated, we wondered whether the traditional linear description of value sensitivity in PFC [14-23] (Figure 6.1C) is an oversimplification.

We hypothesized that ACC and OFC were particularly likely to display more varied tuning to value than the linear functions previously described. Firstly, both regions are sensitive to uncertainty [22,24-28], suggesting that they are plausible candidates for probabilistic population codes in which uncertainty is explicitly represented [2]. The OFC in particular is also implicated in decisions involving confidence [29,30], which is a readout of decisional-uncertainty [31]. Similarly, both regions play a crucial role in value learning. OFC is more specialized for cue-value coding [32-35] , with ACC more concerned with estimating the values of actions, including engage/disengage decisions [36-38]. In both cases, representing current beliefs as probability distributions rather than point estimates brings substantial advantages [31,39], allowing learners to update their beliefs according to their uncertainty [24]. Again, this recommends a probabilistic population coding scheme [1,2]. Finally, the fMRI data presented in the previous chapter hinted at non-monotonic coding in the ACC, whilst previous data from the OFC has suggested that some reward-sensitive cells are most responsive to cues of intermediate value [40,41], as might be expected of a population code for value.

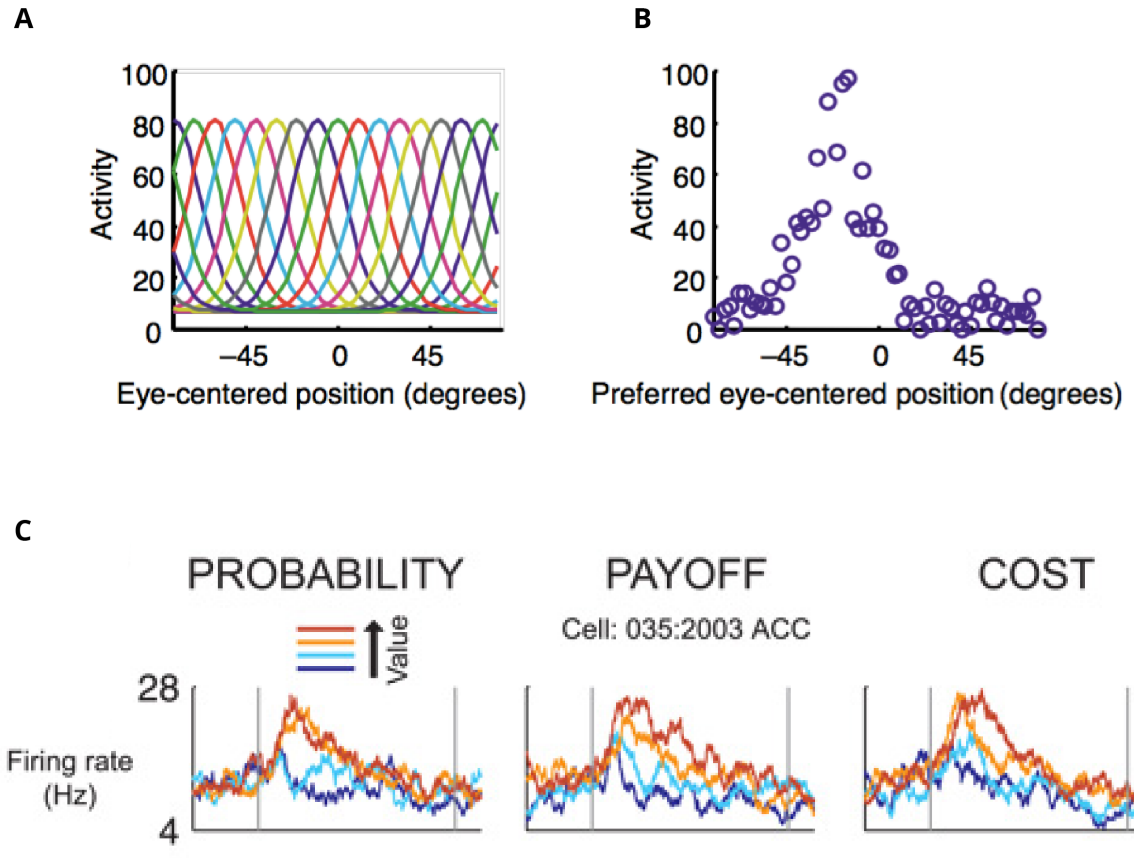


Figure 6.1 | Gaussian tuning for orientation and linear tuning for value (A) Graphical representation of orientation tuning in V1. Each curve is a single cell. **(B)** Evoked activity to a line oriented at 100 in the population of cells represented in panel A. Both adopted from [42]. **(C)** A neuron in ACC responding linearly to cues indicating increasing value across three domains (reward probability, reward size, and cost to obtain reward). Adapted from [15].

We also report data recorded from LPFC and vmPFC, although our predictions regarding these regions were weaker. Analysis in these brain regions was therefore exploratory and reported for completeness. Recent work suggests that single cells in vmPFC might be responsible for value comparison [43] rather than the representation and updating of values, and are associated with a different class of biophysical models [44]. LPFC has been linked with myriad functions, from a well-documented role in support of working memory [45] to a characterization as a domain-general decision region [46]. Neither is as consistently implicated in the learning and updating of value estimates as ACC and OFC, which is the function to which probabilistic codes are most naturally suited.

6.3 Methods

The data discussed here were presented previously in [47], and were collected by Nish Malalasekera, Laurence Hunt, and Steve Kennerley.

6.3.1 Experimental design

Two monkeys (M and F) completed an information gathering task (Figure 6.2A). On each trial, subjects made a decision between a left and a right option, indicating their decision using a joystick. Each option comprised two cues, which corresponded to the magnitude and probability of the juice reward associated with that option. There were 10 cues, corresponding to 5 levels of value each for magnitude (0.15AU, 0.35AU, 0.55AU, 0.75AU, 0.95AU) and probability (10%, 30%, 50%, 70%, 90%) (Figure 6.2B). Preliminary analysis confirmed that animals distinguished between all 5 levels, and treated the two attributes as equivalent [47]. Information was revealed sequentially, and triggered by sustained fixation (Figure 6. 2A). On half of the trials, the first two cues were associated with the same option ('Option trials'). On the other half, the second cue was of the same attribute (magnitude or probability), but for the other option ('Attribute trials').

After the first two cues were fixated, the animal could choose either to gather further information or make a choice. After choice, all information about both options was revealed, before the selected option was highlighted and the corresponding juice reward delivered via oral tube (Figure 6.2A). For the purpose of the current analysis, we focus upon the neural responses to the presentation of the first cue (Figure 6.4).

6.3.2 Neural recordings

Neural activity was recorded simultaneously from ACC, LPFC, OFC, and vmPFC using tungsten electrodes (Figure 6.3). Over multiple sessions, this yielded several hundred cells for each brain region (ACC:198, LPFC: 156, OFC: 195, vmPFC: 160).

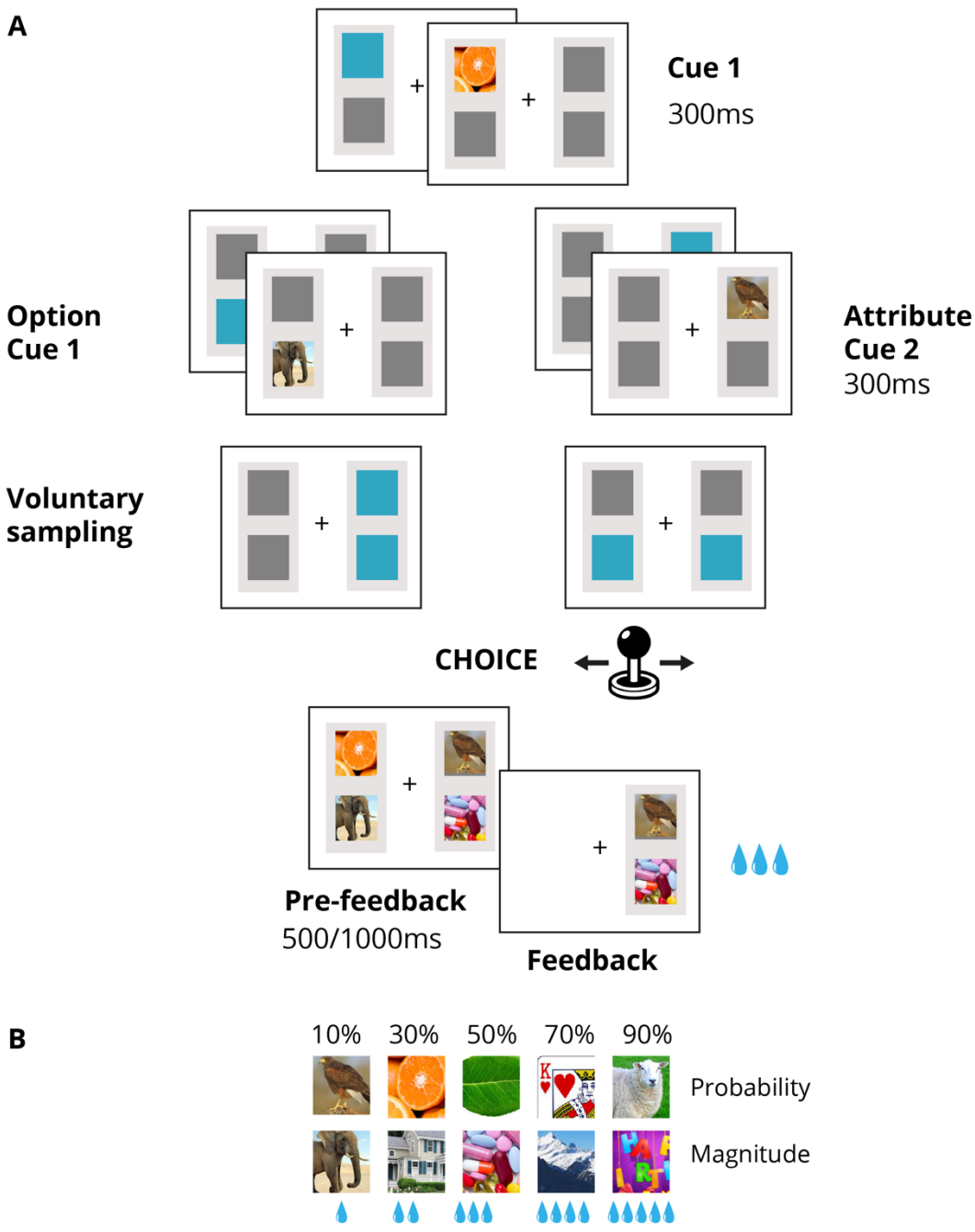


Figure 6.2 | Experimental design (A) Information gathering task. On each trial, monkeys made a left or a right choice based upon sequentially revealed information about the magnitude and probability of juice reward associated with each option. Information was sampled using saccades and the final choice was made using a joystick. See text for details. **(B)** Example picture cues. Each cue provided information about either the probability or magnitude of reward. Adapted from Malalasekera (2015) [47].

For clarity regarding recording location, we quote directly from Malalakasera (2015)[47], in which the data were originally presented:

> *'Neuronal data was recorded from four target regions; ACC, LPFC, OFC, vmPFC. We considered ACC to be the entire dorsal bank of the anterior cingulate sulcus from AP 27-37. LPFC recordings spanned both dorsal and ventral banks of the principal sulcus but were concentrated towards the former. All neurons recorded lateral to the medial orbital sulcus and medial to the lateral orbital sulcus were considered OFC. Finally, vmPFC was considered to be a continuous region which was ventral of the genu of ACC and medial to the medial orbital sulcus. Electrophysiological and depth observations (i.e., gyral and sulcal landmarks, white matter zones) obtained from each electrode during the electrode lowering process were used to estimate the location of each recorded neuron with reference to previously obtained MRI images.'*

Data were recorded at 40KHz. Units were isolated by manual off-line spike sorting (Plexon Offline Spike-Sorter, Dallas).

6.3.3 Analysis of firing rate

We downsampled spiking data to 1KHz, and epoched activity from -500:500ms relative to the onset of Cue 1. In order to calculate firing rate, we used a sliding window of 200ms, with a stride of 10ms, taking the mean inside each window to calculate the firing rate for that bin. This produced a firing rate vector of 81 bins for each trial, with Cue 1 presentation occurring in the 40^{th} bin. For example firing rate plots, see Figure 6.4.

In order to accommodate differences in firing rates across neurons, we normalized firing rate on each trial by that neuron's mean and standard deviation of firing rate. This was computed by smoothing with a 199ms window and then calculating the mean and average firing rate across all time points and all trials.

6.3.4 ANOVAs for value sensitivity

At various points in the analysis we used ANOVAs to test whether a given cell is sensitive to value. This is a departure from the conventional approach for assessing value sensitivity, which uses linear regression, and thus assumes linear tuning functions.

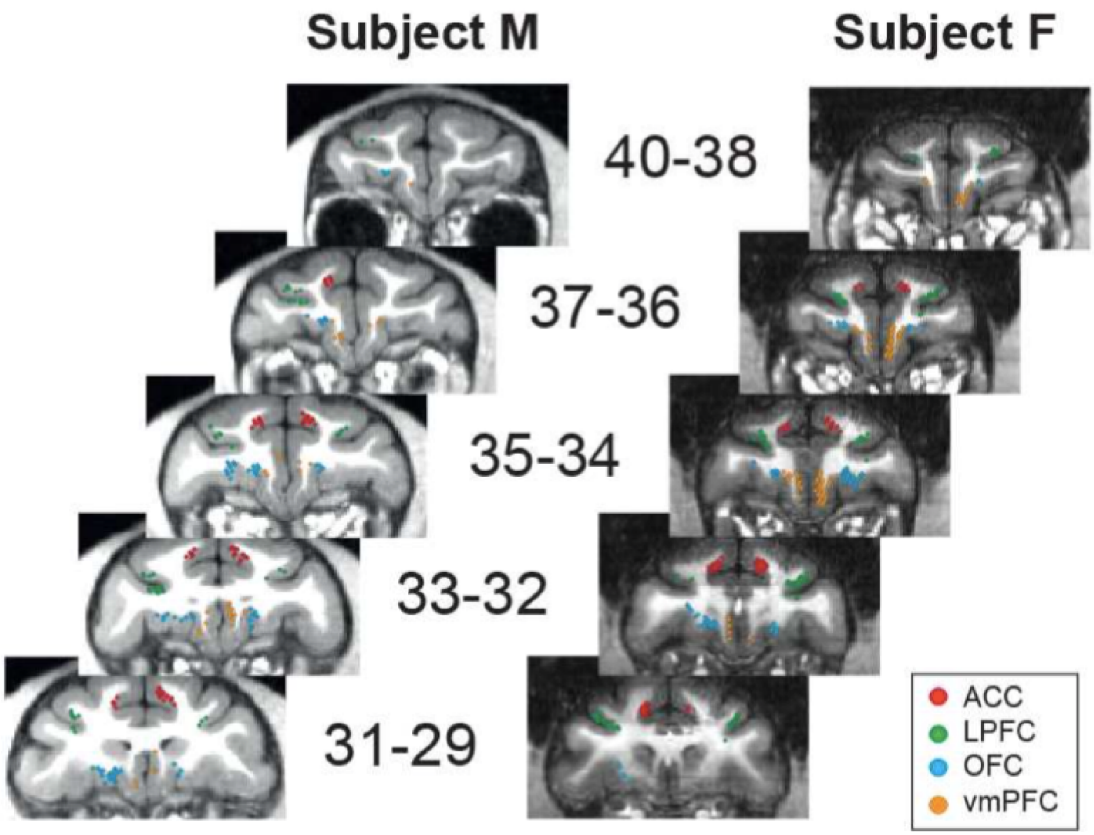


Figure 6.3 | Recording locations Estimated recording locations, with each dot corresponding to a single neuron. Numbers denote millimetres anterior to the Anterior-Posterior line. Adapted from Malalasekera (2015) [47].

We took the mean firing rate on each trial, focusing on an outcome window from 100:500ms. We then used a one-way ANOVA to assess whether variance in firing rate was related to the value of the cue presented on each trial. As a control period, we use the time window prior to cue presentation (-500:0ms).

6.3.5 Binomial testing

In all single-neuron analyses, we used a threshold of p=0.05 to define whether a statistical test was significant for a given neuron. We then tested the fraction of cells in each brain area which surpassed this threshold using a binomial test with an expected probability of 0.05.

6.3.6 Linear and Weibull fitting

In Figure 6.5 we pre-process the data in various ways to remove linear (Figure 6.5A-C), monotonic (Figure 6.5D-F), or attribute-specific (Figure 6.5G-I) components of value. We performed these increasingly strenuous tests to check whether the residuals in each case still retained information about value, allowing us to characterize activity as non-linear, non-monotonic, and attribute-general. In all cases we started with an evoked firing-rate vector, calculated as above, and trial-labels either for value (1-5) [linear and monotonic tests] or all cues (1-10) [attribute residual test].

For linear fitting, we performed a regression, describing Firing Rate (FR) on each trial as a function of a constant, the value of the presented cue, and an error term:

$$FR = \beta_0 + \beta_1 Value + \varepsilon$$

Equation 6.1

We then used the residual vector, ε, as an input to an ANOVA in which we assessed whether the residuals retained information about value.

In order to fit a variety of monotonic functions, we made use of the cumulative Weibull distribution:

$$FR(x) = A(C - 2^{-\left(\frac{x}{T}\right)^k})$$

Equation 6.2

This required non-linear fitting, performed using the matlab function *lsqcurvefit*. We found that this fitting was somewhat temperamental, resulting in fits that varied substantially according to the starting position of the optimization. We therefore initialized the optimization 10 times for each neuron, picking starting parameter values from a normal distribution with mean 0 and standard deviation 1. We then used the best-fitting curve, as assessed by the sum of squared error for each iteration. As with the linear analysis, we then used the residuals from this analysis as an input to an ANOVA testing for traces of non-monotonic value sensitivity in each neuron.

6.3.7 Dot-product analysis of residuals for magnitude and probability

We wanted to check whether the non-linear components of value coding were correlates of value or cue-specific idiosyncrasies. We reasoned that if a cell's linear residuals are truly value coding, then they should look similar when computed separately for each attribute (magnitude or probability) i.e. residuals for each attribute should be correlated.

We repeated the linear regression above but including separate terms for magnitude and probability trials and their associated value. We then took the average residuals (across trials) for each value, for magnitude and probability, resulting in two vectors of length 5. We then took the dot product between the two to assess the similarity of non-linear value representations in each attribute. We compared the true dot-product to a null distribution obtained by shuffling each attribute independently and taking the dot product between the shuffled vectors of 5.

6.3.8 Regression modelling of cue responses

In order to characterize tuning curves, we used a multiple regression to quantify the neural responses to each cue at every timepoint. In analyses in which we focus upon responses to value, irrespective of attribute, we used a nTrials x 5 predictor matrix. Each row contained a single 1, denoting the cue-value presented on that trial. For analyses in which probability and magnitude cues are treated separately, we used a nTrials x 10 matrix, following the same method.

We thus obtain a series of βs, describing each neuron's response to each cue at each timepoint. In Figure 6 and 7 we plot betas from value and an attribute-specific regressions, summarizing each neuron's response by taking the mean β across the outcome period (100:500ms). These β's also form the input to the Gaussian and linear fitting procedures and population decoding procedures (below).

6.3.9 Distance plotting

Following a procedure first outlined in [8], we used sorting by peak response to visualize tuning curves. We used the results of the value ANOVA (see above) to select cells which showed some sensitivity to value. We then took the regression coefficients (βs) derived from the value regression described above and normalized them such that all cells had a minimum β of 0 and a maximum of 1.

In Figure 6.6D,6.6E, and 6.6G, we sort these cells according to their preferred value: the value which received the largest β in the value regression analysis. To produce Figure 6.6D, we binned responses by each value's position relative to the preferred cue. For example, if a neuron preferred value 1, then 2 would be assigned as +1, 3 as +2, 4 as +3, and 5 as +4. A neuron preferring value 3 would be distributed as -2 (value 1), -1 (value 2), 0 (value 3), +1 (value 4), and +2 (value 5). In this manner, we obtain a visualization aligned to each neuron's preferences, where values which are close to the most preferred are central, and increasing distance from the middle corresponds to values that are increasingly distant from the preferred value. In Figures 6.6G we repeat the procedure, but exclude cells preferring either value 5 or value 1, thus removing cells which might be positive or negative linear coders of value. In both cases we use sign-rank tests to assess whether values that are increasingly distant from the mean show decreasing responses.

For Figure 6.6E, we again sort cells by their preferred value. We then simply plot the average tuning curve for all cells preferring that value. The same procedure is used in 6H & I, but we use β's derived from an attribute-specific regression, and assess attribute-general coding by using one attribute to sort the cells (by preference e.g. prefer magnitude=1, prefer magnitude=2 etc.), and then plotting the responses to the other stimulus. This gives a visual approximation to the mutual information between the two tuning curves.

Figure 6.7 combines the approaches described above to produce cross-attribute distance plots. Here we allocate cue responses to bins according to one attribute, and plot the responses to the other.

6.3.10 Comparison of Gaussian and linear fits

In order to formally compare Gaussian and linear characterizations of tuning curves, we fit βs from the cue-value regression with linear:

$$FR(x) = ax + b$$

Equation 6.3

and Gaussian tuning curves:

$$FR(x) = ae^{-\frac{(x-b)^2}{c}}$$

Equation 6.4

where x is value, 1:5.

We assessed the fit for each neuron using Sum Squared Error (SSE). In order to compensate for the higher number of parameters in the Gaussian than the linear model, we compared the difference in fits (SSE_{Linear}-$SSE_{Gaussian}$) to a null distribution of SSE differences composed by shuffling trial labels 1000 times and fitting the resultant βs (Figure 6.8A). We then compared the difference in SSE obtained using the true βs to the distribution obtained using randomised ones (Figure 6.8A iv).

6.3.11 Population decoding

We used Support Vector Machines (SVMs) to estimate population information content in the absence of linear information. SVMs are classifiers which treat examples as points in high-dimensional space, with dimensions defined by the features (the different things which have been measured). In our case, each trial can be considered a point in a high-dimensional space with one dimension per neuron. SVM's attempt to find a 'hyperplane' which maximally separates different classes – in our case, cues of different value [48]. In order to achieve this, an SVM is trained on a series of examples, each of which consists of a class label (value 1:5) and measurements from a set of features (in this case, the activity evoked by that cue in a series of different neurons). Having trained on part of the dataset, the SVM can then be tested on the remaining data, and accuracy calculated by comparing the predicted labels with the true labels in the test set.

In order to obtain trial x feature matrices suitable for analysis with SVM, it was necessary to equalize the number of trials recorded from each neuron for each cue. We set a target number of trials per cue (20). Cells with fewer than 20 trials for any cue were discarded, and neurons-cue combinations with more than 20 trials were clipped to that length, such that only the first 20 trials were retained. This resulted in 197 neurons in ACC, 144 in LPFC, 195 in OFC, and 154 in vmPFC.

We used a linear SVM with a cost parameter of 0.01, tested with 5-fold cross-validation, using the SVMtrain function from LibSVM (available from http://www.csie.ntu.edu.tw/~cjlin/libsvm/)[49]. Accuracy was calculated as the average % correct classification over these 5 folds.

One potential difficulty is that neural populations which contain bountiful information about cue *identity* might give spuriously high classification according to cue *value*. This is clearly an issue with classification analyses, in which the only objective is to find a hyperplane that separates cues of different values. To overcome this, we composed a null-distribution which incorporated cue identity classification, by shuffling the value of attribute and magnitude independently. This produced a situation in which cues of different values were labelled as having the same value (e.g. magnitude=4 and probability=3 are both labelled as value=1). By comparing our true classification accuracy to this null we are thus asking whether the correct pairing of magnitude and attribute cues according to their value produces a superior classification to pairing at random.

6.4 Results

6.4.1 Characterising value-tuning

We observed a wide variety of response patterns to Cue 1, which we summarised by defining tuning curves for each neuron based upon the average response in the period 100-400ms post-cue (see Methods). In Figure 6.4 we preview three distinct response profiles to cues of different values. Neuron 124 in ACC (top row) displays monotonically increasing responses to cues of increasing value, with responses that look similar across both attributes (magnitude and probability). Note that although responses are monotonic, they do not appear to be linear, being sigmoidal in form. Neuron 98 in OFC (middle row) shows positive linear tuning for value with probability cues, but negative linear tuning for value in the magnitude domain. Neuron 123, also in OFC (bottom row) displays a strikingly dissimilar profile: responses are most vigorous for cues of intermediate value, with evoked firing rates dropping off as values become more extreme. Importantly, responses look similar irrespective of attribute, implying that this cell is tuned for value. We next sought quantitative confirmation of substantial non-linear contributions to value-tuning in PFC.

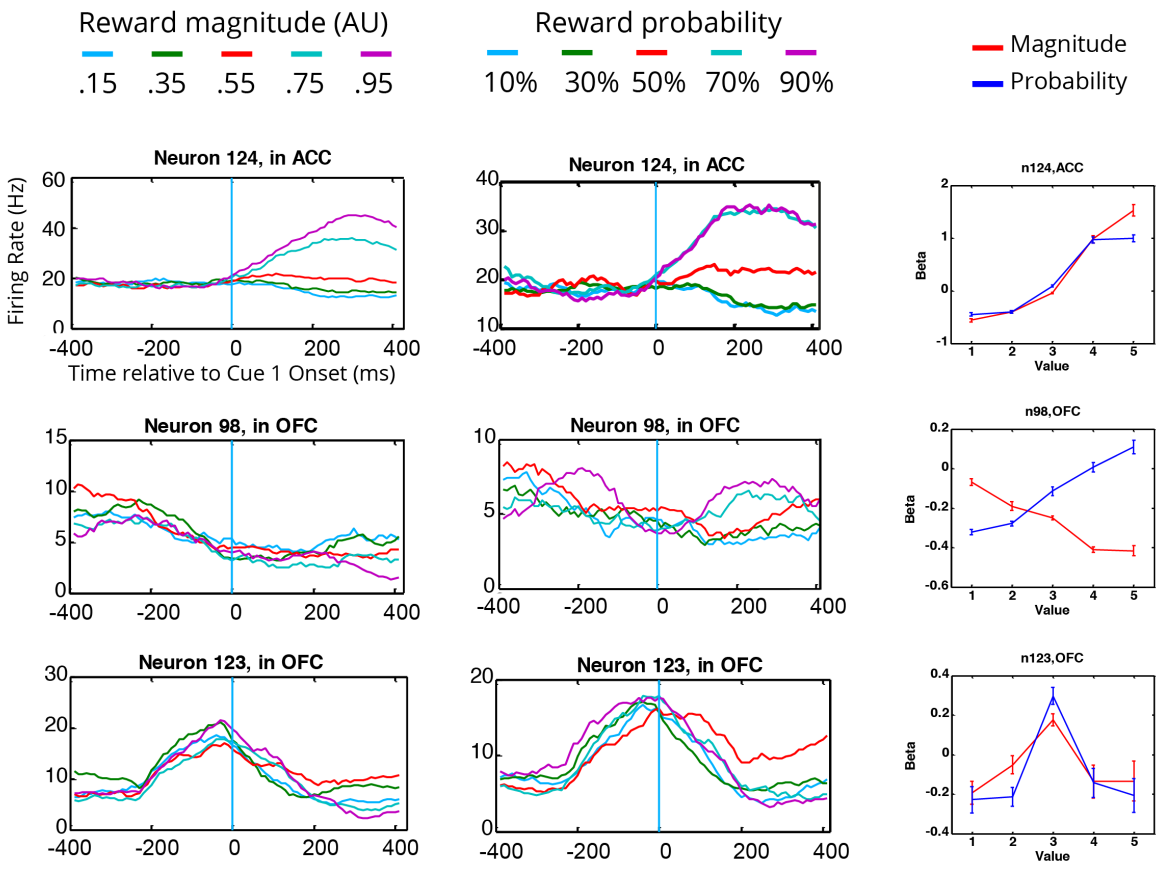


Figure 6.4 | Cue 1 responses Average firing rates for each magnitude cue (left column), and probability cue (middle column). The evoked response can be summarised in a tuning curve, with a single coefficient (β) describing the response to a given cue over the time period 100:400ms post-stimulus (right column). Here we show several different possible tuning functions; linear tuning for both probability and magnitude, which might be described as value-tuning (top right); anti-correlated linear tuning for probability and magnitude (middle right); and non-monotonic tuning for both magnitude and probability (bottom right). Error bars in tuning-curve plots are SEM of betas across outcome period.

6.4.2 PFC populations retain value-selectivity after linear regression

Probabilistic population codes support gain-encoding of uncertainty if codes are non-linear (such as densely sigmoid or Gaussian) [2,3]. Our first set of quantitative analyses (Figure 4.5) tested for this non-linear component, by removing linear and monotonic value signals, and assessing whether the resulting representation still contained information about Cue 1 value.

We first summarised the response of each neuron on each trial by taking the average firing rate for the outcome period (100-400ms). We then estimated a linear regression model describing the relationship between value (1:5) and responses for each neuron on each trial (Figure 6.5A). We then took the residuals from this analysis (Figure 6.5B) and subjected them to an ANOVA, asking whether the residuals still contained information about the value of the cue presented on each trial. As control, we performed exactly the same analysis upon activity estimated pre-Cue (-400:0ms).

We found that in ACC, LPFC, and OFC, a larger-than-chance population of neurons retained non-linear selectivity (all $p<0.001$, binomial test) (Figure 6.5C). This was particularly prominent in OFC, where almost 30% of cells still carried information about cue-value after the removal of linear components. The pre-cue analysis confirmed that <5% of cells were significant using control data (Figure 6.5C).

We next asked whether value-selectivity was retained after removing all *monotonic* components of value coding. This is important because slight deviations from linearity might arise through subtle misattribution of value in the learning process, perceptual biases, or the biophysical factors. The non-linearity result above tells us that coding is not perfectly linear; non-monotonicity, however, implies a different coding scheme entirely.

We therefore repeated the model-estimation and residual analysis above, but using non-linear fitting of Weibull functions in place of linear regression [50] (see Methods). We fit a Weibull to average firing rate for each value (Figure 6.5D), again generating a set of residuals (Figure 5E), which we then subjected to an ANOVA. Residuals from Weibull fits in OFC ($p<0.001$, binomial test) and LPFC ($p=0.028$, binomial test) retained value selectivity, whilst those for ACC ($p=0.19$, binomial test) and vmPFC ($p=1$) were no longer selective having removed monotonic components (Figure 6.5E). This suggests that non-linear tuning in ACC and LPFC is likely still monotonic. As before, the pre-cue analysis confirmed that <5% of cells were significant using control data (Figure 6.5E).

One challenge associated with the present dataset is that each value is associated with only two cues, which raises the possibility than non-linear coding might arise from cue, rather than value, selectivity. If, for instance, a given neuron has a preference for a particular picture, this will result

in a bump in the tuning curve which is poorly described by a linear fit. To guard against this possibility, we next repeated the linear fitting procedure separately for probability and magnitude. We then asked whether the dot product of the residuals in the two domains (Figure 6.5G) was greater than we would expect by chance, as assessed by a permutation test where betas for probability and magnitude were independently shuffled (Figure 6.5H).

We found that residuals for magnitude and probability were correlated in ACC ($p=0.024$), binomial test), but in none of the other brain regions (LPFC, $p=0.137$; OFC, $p=0.42$; vmPFC, $p=0.36$) (Figure 6.5I). This suggests that deviations from linearity in ACC do reflect value coding, and argue against a cue-coding explanation for non-linear selectivity. The failing to find the same in OFC and LPFC suggests that non-linear components of value coding in these regions might be related to cue coding.

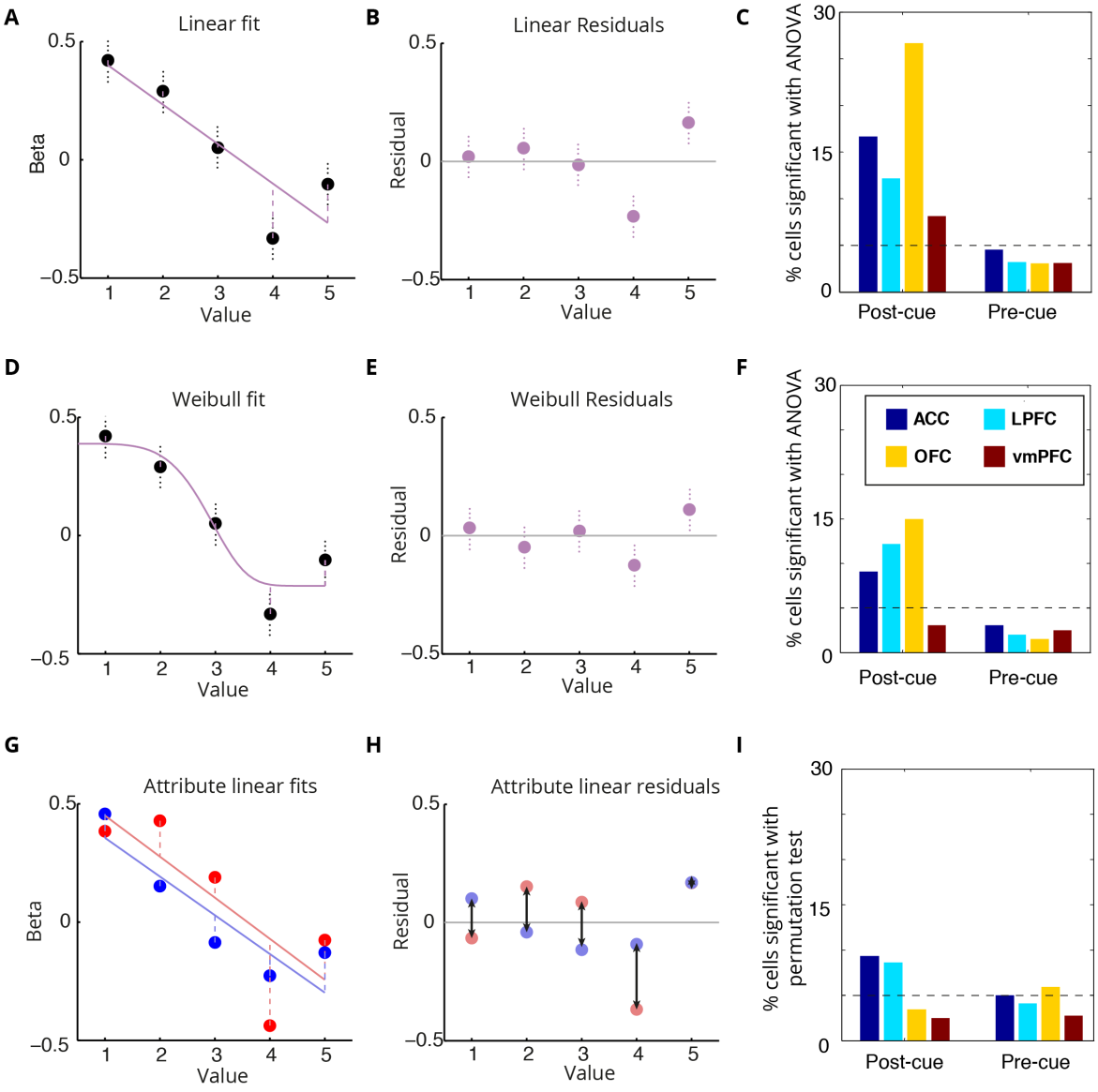


Figure 6.5 | Removing linear and monotonic value representations Panels ABDEGH demonstrate the methodology on a single neuron, whilst CFI are population summaries. We start by quantifying the response of a neuron to cues of value 1-5. We can then remove linear components of value **(A)**, leaving us with residuals **(B)**. For each neuron, we can then ask whether the residuals still discriminate between cues of different value. In all brain areas, a significant portion of cells retain selectivity after removing linear value components. Pre-cue analyses confirmed that no such selectivity was observed in the pre-cue period. **(C)**. Going one step further, we can fit a Weibull to accommodate non-linear, monotonic tuning curves **(D)**. The residuals from this process **(E)** also retain value information in OFC, and LPFC, but not in the pre-cue period **(F)**. To guard against the possibility that residuals are capturing single-cue selectivity, we can check whether linear residuals for magnitude and probability **(G)** are correlated **(H)**. By comparison to a null distribution composed of betas derived from random permutations of

trials **(I)**, we find that non-linear cue representations share structure in ACC (p=0.0244), suggesting that non-linear components do indeed reflect non-linear value-tuning rather than cue selectivity. Again, pre-cue analysis confirmed that no brain regions showed above-chance correlation before cue presentation.

6.4.3 Visualizing non-monotonic tuning curves

Plotting tuning functions for hundreds of neurons provides a challenge. A seminal paper describing tuning curves for number in the PFC [8] provided several useful visualisations (Figures 6.6A-C), which we attempted to replicate. The first (Figure 6.6A) involves sorting cells by their preferred number and plotting their response to other cues as an increasing distance from the peak. By testing adjacent responses against one another, we can assess whether responses decrease as we get further away from the peak, as predicted from Gaussian tuning. Secondly, using the same sorting of cells by preferred number, one can plot average normalized tuning curves corresponding to cells with the same preferred number (Figure 6.6B). Finally, we can simply plot a histogram of preferred number over cells (Figure 6. 6C).

Plotting a distance visualisation of response profiles for value-selective cells (classified by ANOVA) reveals a striking 'pile' appearance very reminiscent of the original [8] (Figure 6.6D). We tested for differences between adjacent values (excluding the comparison between peak and adjacent, which is biased by peak selection), and found that all but one of the differences (+3/+4) was significant (sign-rank tests, Z between 2.41-13.32, p from 0-0.0157). Plotting tuning curves sorted by preference (6.6E) and a histogram of preferred values (6.6F) also provided clear indications of non-monotonic tuning in the dataset. Interestingly, we observed similar distributions of preferences over values as previously reported over numbers [8], with larger populations of cells preferring end values than intermediate ones (Figure 6.6C and 6.6F).

Although these plots are visually compelling and have previously been used to evince Gaussian tuning [8,9], we were concerned about two possible confounds, arising from the contribution of monotonic and cue-specific coding.

Firstly, distance analyses can be generated from monotonically tuned neurons alone. If there are both positively and negatively tuned neurons in the population, both of the sides of the distance plot will be monotonically decreasing as we move away from the preferred value. The 'pile' could thus be composed from two juxtaposed linear tuning curves, one negative and the other

positive. To ensure this was not the case in our dataset, we repeated the distance analysis having excluded end-preferring neurons (those with preference for value=1 or value=5). Both the appearance and the significance of the analysis was largely preserved (p between 0.0268 and 0.04), with the comparison of +2 to +3 at trend level (p=0.10).

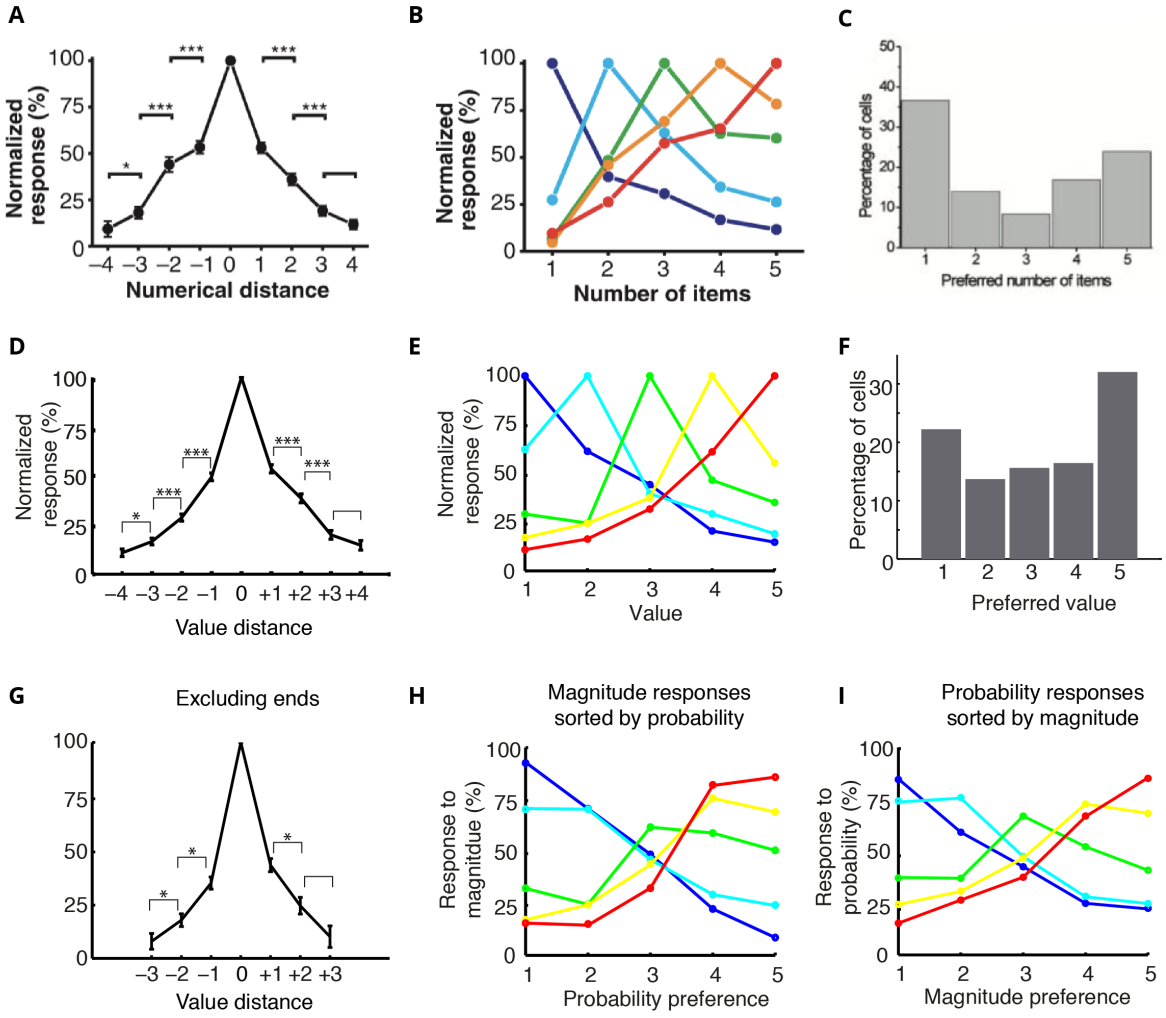


Figure 6.6 | Visualizing tuning curves (A) Nieder et al [8], plot PFC responses to cues of varying number as function of distance from the number eliciting maximal activity. The drop off of response with increasing distance from preferred number implies Gaussian tuning. (**B)** Tuning curves for cells sorted by number preference and **(C)** distribution of number preferences in PFC, again from [8]. We performed the same set of analyses to assess evidence for non-monotonic value-tuning. Both our distance **(D)** and tuning **(E)** plots are similar to those found by Nieder et al. The distribution of preferences over cells **(F)** is biased towards cells preferring 5, whereas Nieder et al find a bias towards cells preferring 1. **(G)** To ensure that the significance of our

distance analyses was not due to the presence of monotonic-coding cells, we excluded cells preferring cues 1 and 5 from the analysis. The resultant distance plot is similar, with statistical significance unaffected. Furthermore, sorting tuning curves for value-selecting cells based upon probability preference and plotting responses to magnitude **(H),** or vice versa **(I),** produces well-ordered curves. This generalization of tuning between attributes is predicted if cells are representing value, as opposed to single attribute or cue coding.

Secondly, we wanted to confirm that tuning curves are not attribute specific. If the tuning curves we visualize here correspond to non-monotonic representations of value, we ought to be able to predict tuning curves for magnitude from those for probability, and vice versa. To test this, we sorted by magnitude or probability and plotted tuning curves for the other attribute (Figure 6.6H & I). On average across brain areas, we found that tuning curves were preserved across attribute, arguing against an attribute or cue-specific interpretation.

Closer inspection of distance plots for each brain area without end-coders (Figure 6.7) confirmed that cross-attribute tuning was present in all brain areas. However, there was some variation, with tuning clearest in ACC and LPFC and more disordered in OFC and vmPFC. In all brain areas, the highest average response to magnitude cues was predicted by the peak response to probability, and vice versa.

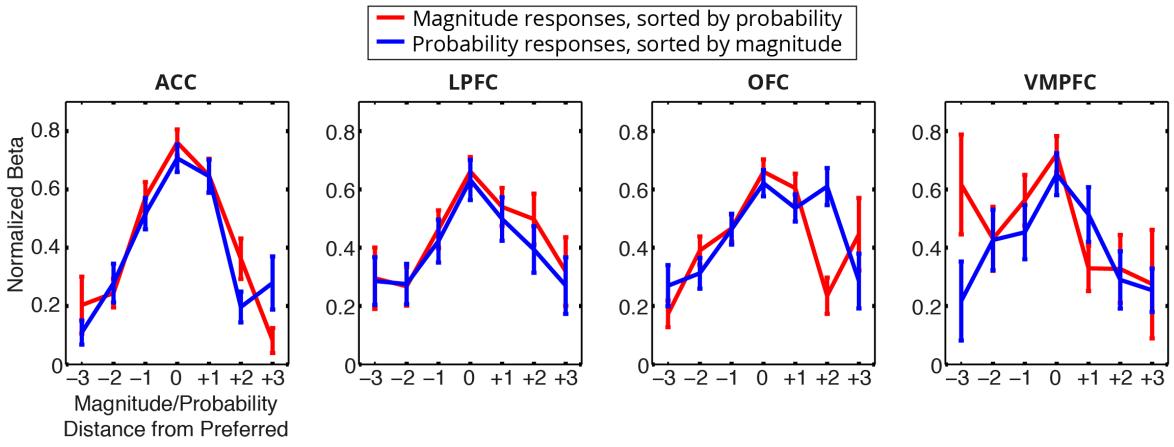


Figure 6.7 | Cross-attribute distance plots by area Segregating cells by area and repeating the distance plots, excluding end coders (as in 6.6G). All areas show a degree of cross-attribute tuning, although it is more pronounced in ACC and LPFC than OFC and vmPFC.

6.4.4 Testing Gaussian vs. linear coding

So far, we have seen that neurons in PFC retain information about value once linear and monotonic information has been stripped away, and that population visualizations resemble those of Gaussian-tuned neurons. In our next set of analyses we provide more direct evidence for this claim, by explicitly comparing the ability of Gaussian and linear functions to capture tuning curves for value. This approach has the distinct advantage of not requiring us to filter or censor the data in any way. Thus, we allow linear trends to be accommodated in the shape of Gaussians (for example, neuron 125 in Figure 6.8B), simply asking which model has a higher probability given the data.

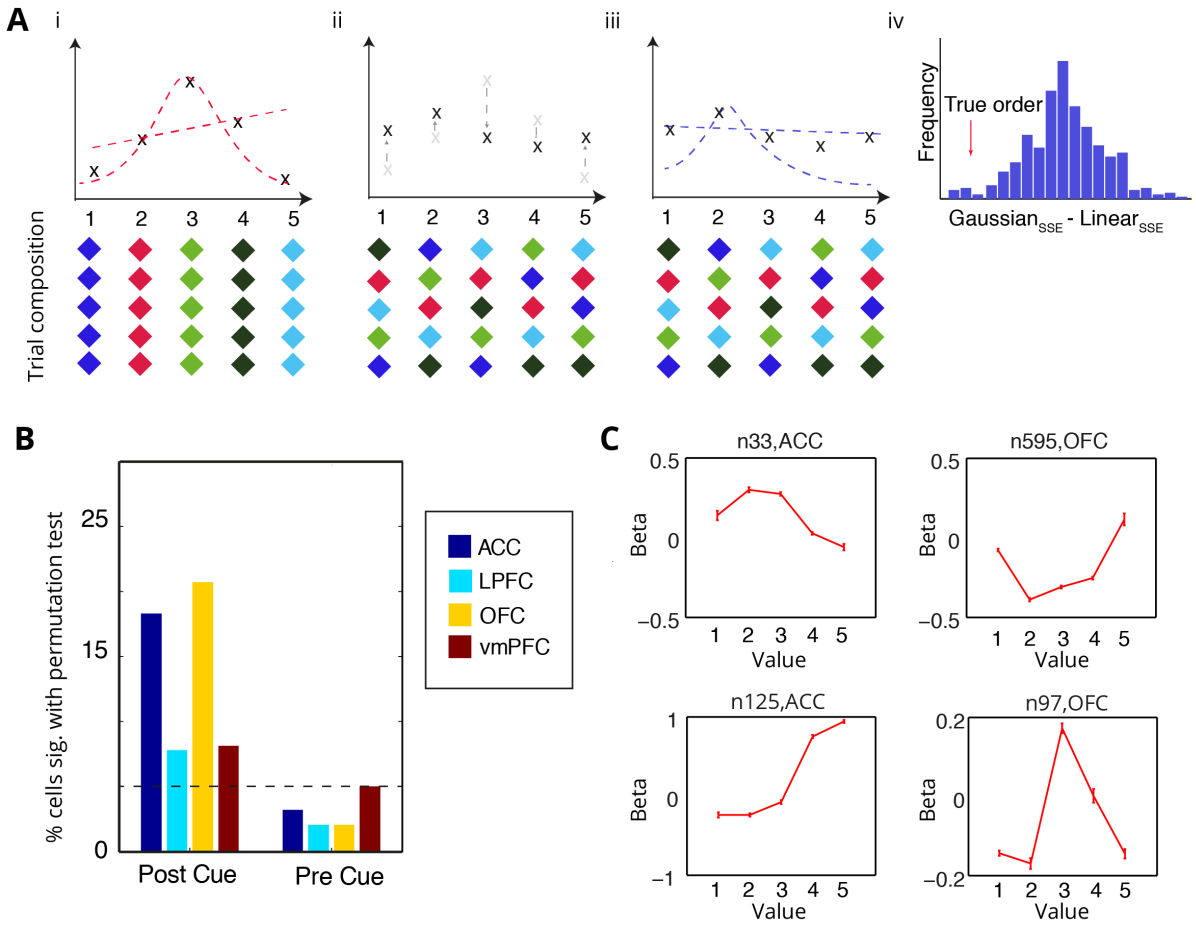


Figure 6.8 | ACC and OFC contain Gaussian-tuned neurons (A) We used permutation testing to assess whether neurons were better fit by linear (2 parameters) or Gaussian (3 parameters) tuning functions. (i) Fitting was performed upon betas derived from a multiple regression, summarizing each neuron's response to each cue. We fit a single tuning curve across both value

domains (magnitude and probability). For each neuron, we shuffled trials (ii) and refit our tuning curves each time (iii). This provided a null distribution of differences in Sum Squared Error (iii), to which the true SSE difference was compared. **(B)** ACC and OFC contained sizeable populations of Gaussian-tuned neurons. **(C)** Example neurons. Notice that this approach recovers cells that are 'classically' Gaussian (n33 and n97), as well as negatively tuned (n595) and non-linearly tuned (n125) neurons.

To this end, we used multiple regression to summarise each neuron's responses to the 5 cue-values, and then fit linear and Gaussian tuning functions to the resultant regression coefficients. We used a permutation-test to assess whether the difference between the Gaussian and the linear fit was greater than we would expect by chance (Figure 6.8A). This involved random shuffling of trial labels, following which the same procedure was repeated, producing a set of βs (Figure 6.8Aii), which we then fit with Gaussian and linear functions (Figure 6.8Aiii). We then compared the difference in SSE obtained using the true βs to the distribution obtained using randomised ones (Figure 6.8Aiv).

ACC and OFC contained substantial populations of cells (18.7 and 20.1% respectively, binomial test both $p<0.001$) whose tuning was better described as Gaussian than linear (Figure 6.8B). Several Gaussian-tuned neurons selected by this method are plotted in 6.8C. Our method detects both classical Gaussian profiles (n33, n97), whilst also picking up non-linear tuning that is well described as one-tail of a Gaussian (n125). Furthermore, we also identify cells that are negatively tuned to value (n595).

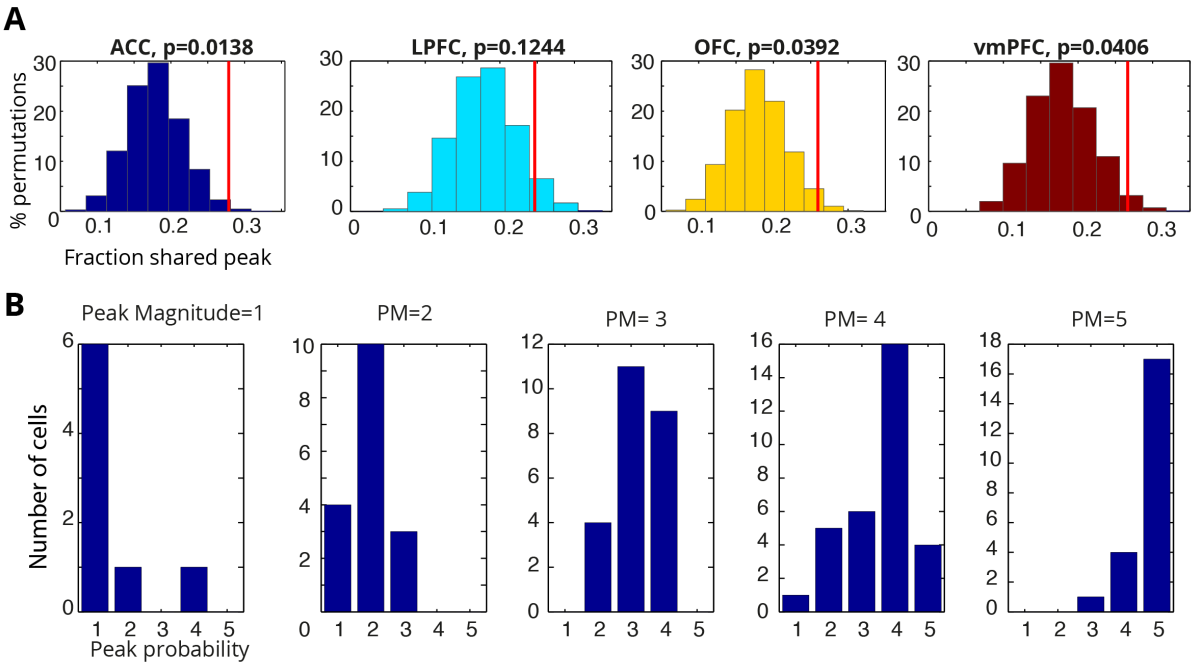


Figure 6.9 | Control analysis: correlating probability and magnitude tuning Discarding cells which preferred either the maximum or minimum magnitude (i.e. monotonic coders), we asked whether preferences magnitude predicted those in probability, as would be expected in cells tuned to value. We calculated the fraction of the population in which the preferences matched, and compared to a permutation test in which the order of cells was randomly shuffled. **(A)** ACC, OFC, and vmPFC all show a significant concordance of non-monotonic magnitude and probability preferences. **(B)** Conditional histograms for all cells classified as Gaussian by the previous analysis. Panels 1-5 correspond to cells with peak response to magnitudes 1-5. Clear peaks in the corresponding probability preference imply value tuning.

6.4.5 Confirming value-coding

In order to reassure ourselves that the Gaussian-tuning identified in the previous section did not correspond to cue-specific responses, we assessed in more detail the degree of concordance between probability and magnitude tuning. We sorted cells by their preference in the magnitude domain, discarded those preferring 1 or 5, and then asked what percentage of cells had a matching peak in the probability domain. We then compared this fraction to a permutation test in which the order of cues was randomly shuffled. In ACC (p=0.0138), OFC (p=0.0392), and vmPFC (p=0.0406), shared-peaks were more common than expected by chance, with LPFC showing a trend in the same direction (p=0.12) (Figure 6.9A). This implies that cells which prefer magnitude cue value 2 are more likely to prefer probability cue 2 than those that prefer

magnitude cue 3 or 4. To reinforce this point, we plotted conditional histograms for those cells identified as Gaussian-tuned in the previous analysis (Figure 6.9B), emphasising that non-monotonic preferences are indeed shared across-attribute, and echoing the distance plots in Figure 6.7.

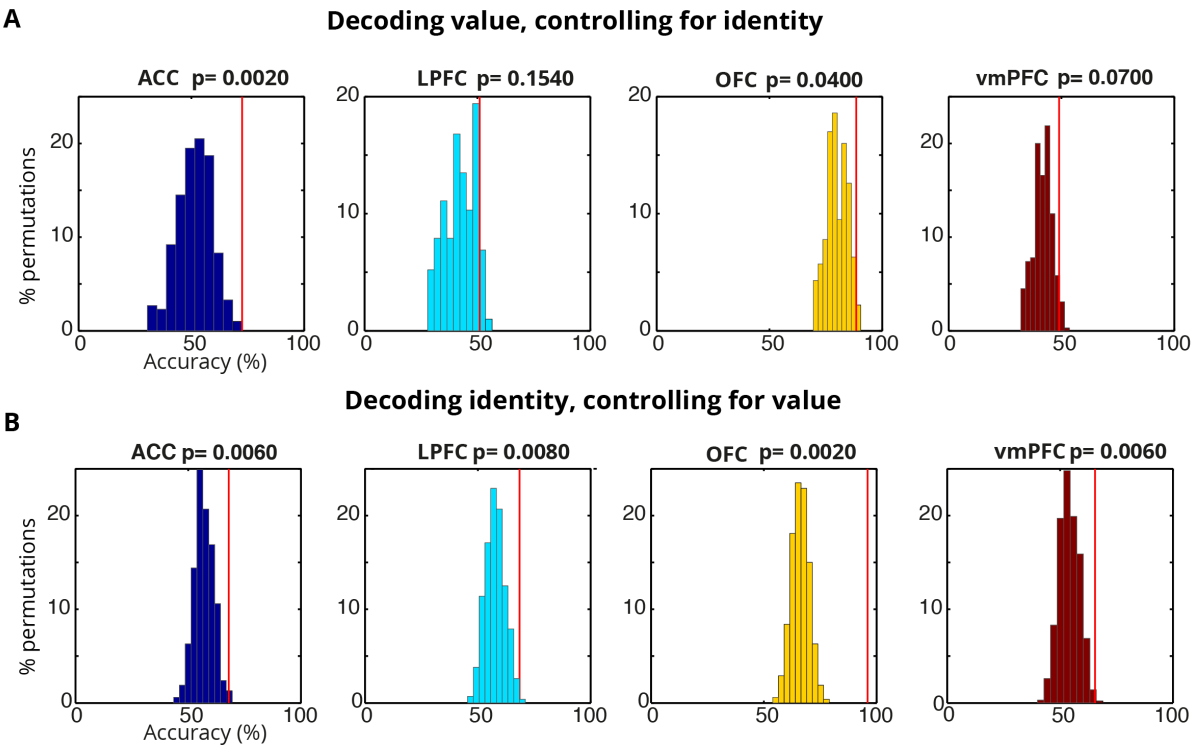


Figure 6.10 | Population decoding of value after linear information removal We used Support Vector Machines to assess the information content of pseudo-populations of neurons from each brain region, having removed linear value correlates via regression. Red vertical lines denote real performance, histograms are performance of random shuffles. **(A)** Cue value is decodable from neural populations in ACC and OFC. The null distribution is composed of cue-shuffles for probability and magnitude independently, such that mere cue selectivity is accommodated by the null (hence high null-decoding in OFC). **(B)** We performed a complementary analysis in which we matched cues for value, and tried to decode cue identity. All brain areas showed successful decoding, with representation of cue identity in OFC markedly stronger than elsewhere.

6.4.6 Population representations of value in the absence of linear information

The analyses above provide evidence that some single cells in the PFC, predominantly in ACC and OFC, possess non-monotonic tuning functions for value that are approximately Gaussian. However, decisions are likely formed through the co-ordinated action of populations of cells in each of these regions, rather than relying upon the representation of value in any one neuron. To this end, we turned to population decoding methods to assess the contribution of non-linear coding to value representations across hundreds of neurons.

For each brain region, we assembled a 'pseudo-population' of cells, comprising all of the cells recorded in that region over all sessions. Such pseudo-populations do not represent the full richness of a true neural population, because cells were recorded at different times. Hence information contained in intercell correlations or temporal dynamics is lost[13]. Analyses of such pseudo-populations thus provide a useful lower-bound on the information represented in a given brain region.

We asked whether value was decodable from each pseudo-population, having removed linear value codes. For each brain region we took all cells in which there were more than 20 trials for each cue, and regressed out linear value correlates. This resulted in pseudo-populations of 197 neurons in ACC, 144 in LPFC, 195 in OFC, and 154 in vmPFC. We then trained a Support Vector Machine (SVM) to distinguish between cues of different value, collapsing across attribute, using the LibSVM package [49]. We compared performance to a permutation test in which the betas for magnitude and probability were independently shuffled, such that cues of different values (e.g. magnitude = 1, probability = 3) were assigned to the same value. This produced a null in which cue-selectivity was accommodated, preventing false positives due to cue-distinguishability. Both ACC ($p=0.0020$) and OFC ($p=0.040$) supported value-decoding in the absence of linear information (Figure 6.10). vmPFC was marginally significant ($p=0.070$), whilst LPFC ($p=0.154$) was not. Both ACC and OFC, therefore, contain non-linear population representations of value, in accordance with the analysis of single-cells performed above.

By way of comparison, we also asked to what extent cue identity was encoded in each population, removing variability due to value. To do this we matched cues from each value level (magnitude 1 vs. probability 1, magnitude 2 vs. probability 2 etc.) and trained an SVM to

distinguish between them. Performance was then calculated by averaging over all values. A population only encoding cue value would thus be at chance in this analysis, because value 1 as indicated by cue 1 would be identical to value 1 as indicated by cue 6. We compared decoding of cue identity to random trial labels. Cue decoding performance in OFC far exceeded other brain regions, standing at an impressive 96% (p=0.002), compared with ACC=68.0% (p=.006), LPFC=68.5% (p=0.008), and vmPFC=65.5% (p=0.006). This suggests that cue identity coding is particularly prevalent in OFC, consistent with the idea that OFC maintains a representation of state space [51,52].

6.5 Discussion

Neural coding in numerous regions is compatible with the use of Probabilistic Population Codes [1,2,31,39,53] to represent and manipulate uncertain information. Recent data [54,55] suggests that such codes might be used to make decisions about orientation and movement of visual stimuli. Given the established role of the prefrontal cortices in representing malleable and uncertainty-modulated value estimates [28], we reasoned that population codes might provide a useful format for computations in value-based choice. To test this idea, we focused on a key predicate of PPCs, non-linear tuning. Using a variety of approaches, we found evidence of non-linear value tuning in the ACC and the OFC.

6.5.1 Value tuning in the anterior cingulate

As outlined in the previous chapter, the ACC plays a central role in the integration of information relevant for value-based choice[28]. The ACC maintains representations of value over a variety of timescales [56], and lesions abolish credit assignment in extended sequences of actions [38], suggesting that value representations are diverse and malleable in this part of the brain. Uncertainty is also known to modulate activity in the cingulate [24,57], with interindividual variability in the speed of belief updates linked to the strength of cingulate representations. The ACC thus appears a prime candidate for probabilistic population coding.

We found abundant evidence that neurons in ACC code value in a non-linear fashion. ACC representations of value were resilient to the removal of linear information (Figure 6.5), with aggregate tuning curves resembling non-linear coders (Figure 6.7). A substantial number of neurons had tuning curves better fit by Gaussians than by linear models (Figure 6.8), and value

could be convincingly decoded from the ACC population having removed linear information (Figure 6.10).

However, when we removed all monotonic correlates of value – thus accommodating sigmoidal, concave, and convex tuning curves - residuals in the ACC were no longer value coding. This suggests that the non-linear tuning observed in ACC is probably monotonic i.e. neurons only ever increase (or decrease) their firing rate with increasing value. Whilst this representation is still consistent with a probabilistic population code format [2], its interpretation requires care.

Firstly, non-linearity might arise due to errors in the calculation of value, either due to imperfect learning by the animal or imperfect estimation of subjective value by the experimenter. Given the preponderance of linear value correlates in other brain regions, this seems unlikely. Similarly, biophysical explanations for this phenomenon seem unlikely given its absence in, for instance, vmPFC.

Secondly, and more problematically, non-linear, monotonic representations of value might correspond to linear representations of a value-related variable. For instance, some regions of frontal cortex appear to represent the best choice given the current evidence [58]. Given the ACC's well-documented (albeit controversial) role in foraging [36,59-61], it is plausible that ACC neurons carry a signal about whether an option is better or worse than the average. This would produce sigmoidal tuning curves, with high value (4 & 5) and low value (1 & 2) coded similarly. Neuron 125 (Figure 6.7B) appears to have such a tuning curve. Viewed through the lens of value, this representation would appear non-linear, and might be fitted better by a Gaussian than by a straight line. This touches upon a deeper philosophical issue to which we will return in due course, concerning our ability to make statements about representations and their form. To preview it here: a non-linear representation of the original attribute (value) is indistinguishable from a linear representation of a non-linear function of the original attribute (such as a choice probability).

6.5.2 Value tuning in the OFC

Recent characterizations of the OFC have focused upon its role in representing the state space associated with a choice [52,62]. However, the rich literature on value-coding in the OFC suggests that this map of task space is also imbued with information about the value of those states [63,64].

Our data suggest that both of these representations are in part non-linear. Both the state space (cue identity) and value are decodable from the non-linear components of neural activity (Figure 6.10).

All of our analyses suggested a substantial non-linear component to OFC value representations. Moreover, representations of value in some cells in OFC are non-monotonic (Figure 6.5E), surviving the removal of a fitted Weibull function. The OFC hosted the largest number of Gaussian-tuned cells (20.1%, or 39 cells) of any brain region, and showed the highest retention of selectivity after the removal of linear information. Our analyses add to the growing appreciation that the OFC representation of value is richer and more diverse than previously appreciated [65,66].

The entanglement of state and value representations in this task makes it difficult to conclusively fractionate the two. Each value was associated with only two cue identities, meaning that cue-identity coding might easily be misinterpreted as value sensitivity (and vice versa). To address this in our analysis of residuals, we compared the linear residuals for magnitude and probability, calculated separately (Figure 6.5I). In the OFC, these residuals were not significantly correlated in a greater number of cells than expected by chance. This implied that non-linear OFC information might be unique to each domain, suggesting coding of cue identity. However, this is in conflict with our distance analyses (Fig 6.6 & 6.7), the correlation in peak tuning across attributes (Figure 6.9) and the ability of SVMs to decode value in OFC when controlling for cue identity (Figure 6.10), all of which point to value, and not state, representations.

Where does this this discord come from? We note that correlations across 5 data points, which is what the cross-attribute residual test amounts to, are ambitious. Furthermore, the small range of values considered here means that only tuning curves centred on the middle value will be unaffected by removal of linear information; neurons preferring values 2 or 4 will be substantially altered. Both of these problems might be addressed by using a larger range of values, in association with a larger number of cues. Even more compelling would be to record data during learning, holding cue identity constant and varying value over time. Recent work in the DLPFC found that single neurons preferentially represent the fluctuating value associated with a specific stimulus [67], suggesting one flavour of identity-value multiplexing.

6.5.3 LPFC and vmPFC

As anticipated, evidence for non-linear tuning in LPFC and vmPFC was comparatively weaker, although not absent. LPFC retained information about value after linear regression, but this was not decodable at a population level when controlling for cue identity. This, and the fact that peak value preference in magnitude and probability were not correlated, suggests that LPFC representations are largely cue-specific. Interestingly, the converse was true for vmPFC, which did not harbour single-cell representations of value after regression but showed a trend towards population-level representations ($p=0.07$), as well as a correlation between tuning for probability and magnitude.

6.5.4 Conclusions: PPC's in PFC?

We found that neurons in ACC and OFC represent value in a non-linear format compatible with probabilistic population coding. Certainly, then, these regions *might* support a probabilistic population code, consistent with their roles in the accumulation of uncertain information and the representation of state space.

However, the identifying feature of such codes is the representation of uncertainty, of which there is none in the current task. This limits our ability to test a key prediction of the model, namely that the gain of population responses fluctuate inversely with uncertainty [2]. Studying the responses of these populations whilst animals learn the values associated with particular cues would allow us to test these ideas more directly, whilst obviating the identity-value confounds we struggled with here. Such data would also allow us to test other models of uncertainty-coding, such as those based upon spike-rate variability over time [68]. This would also allow us to link the responses we observe to the behaviour of the animal, arguably a critical step in characterizing a neural representation [69].

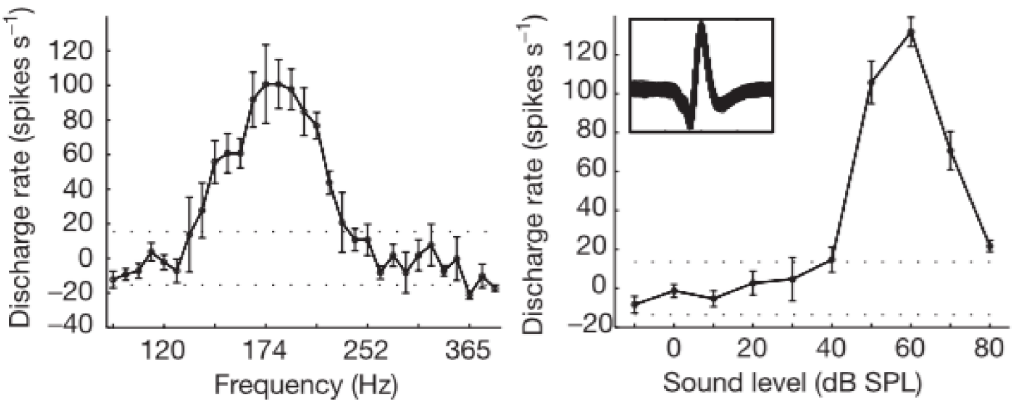


Figure 6.11 | Neural coding of pitch and volume From Bendor & Wang (2006). Cells in the auditory cortex of the marmoset are tuned to frequency and sound level. Conceptually, both variables are continuous and linear, but their neural representation is tuned and non-linear.

6.5.5 What tuning curves should we *expect* for value?

Our results emphasise that intuitively linear concepts need not have a linear neural instantiation. The overwhelming convention in the neuroscience of value-based choice is to assume that variables that go up and down are represented by neurons that fire more or less[14-23], an assumption also implicit in fMRI studies [70-77]. A simple consideration of sensory systems suggests that this is an unhelpful supposition. Pitch, for instance, is a linear concept coded by populations of tuned neurons, as is loudness [78] (Figure 6.11). Recent progress in understanding the neural code for the representation of space has emphasised how efficient neural codes can assume emphatically unintuitive forms [79,80]. Grid cells represent space in a repeating hexagonal code which bears no semblance of a linear relationship with external frames of reference. As models from sensory processing and decision-making become more popular in the study of value-based choice [81,82], we hope that the assumption of linearity falls by the wayside.

A second conceptual distinction is between single-cell and population codes. As our ability to simultaneously measure the activity of many cells improves, it becomes increasingly hard to ignore the fact that population of neurons behave in complex ways that are completely inscrutable to single-neuron analysis [12,13,83]. Grasping that nettle, we must acknowledge that tuning functions that seem sensible for a single cell – such as a linear code for value – don't

seem so effective if they are implemented by hundreds of thousands of neurons simultaneously. What seems sensible for one may be ludicrous for a thousand neurons. Grid cells again provide a salutatory lesson here: the activity of any one cell is weakly informative, but a population organized into modules provides a representational code of quite striking power [84]. Similarly, the non-linear aspects of value-coding that we document here may seem unpalatable when considering the information imparted by the response of a single neuron, but functional and flexible when considered as a single thread in the fabric of a population response, such as that envisaged by PPC's. To quote one of the seminal papers in the field, **'**An entire collection of cells is a terrible thing to waste on representing just a single value of some quantity' [1].

Our understanding of coding schemes in other regions of the brain has been deepened by considering the optimal representation of information from a theoretical perspective [85]. Such information-theoretic approaches have stressed that tuning schemes ought to reflect the natural statistics of the environment, capturing the principle axes of variation whilst discarding redundant information [86]. Convergent evidence suggests that aspects of the visual system obey such laws, with synthetic neural networks trained to perform high-level visual tasks frequently recapitulating aspects of tuning in V1 [87]. Recent work suggests that the grid-like representation of space may emerge simply from a convenient way to encode positions in a two-dimensional plane [80,88,89]. Applying such 'efficient-coding' approaches to value representations presents (at least) two challenges [90]. Firstly, we do not have a good understanding of the natural statistics of rewards in the environment, although recent foraging approaches have attempted to recapture some of the ethological validity arguably lost in neuroeconomic approaches. Secondly, value is inherently subjective, unlike the properties of visual scenes. It is a property of not only of the environment, but of the state of the organism in question. Estimating the natural statistics of value may therefore require us to estimate the natural statistics of physiological needs.

6.6 References

1. Zemel, R. S., Dayan, P. & Pouget, A. Probabilistic interpretation of population codes. *Neural Comput* **10,** 403–430 (1998).
2. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nature Neuroscience* **9,** 1432–1438 (2006).
3. Beck, J., Ma, W. J., Latham, P. E. & Pouget, A. Probabilistic population codes and the exponential family of distributions. *Progress in brain research* **165,** 509–519 (2007).
4. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*

(Lond.) **148,** 574–591 (1959).
5. Albright, T. D. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* **52,** 1106–1130 (1984).
6. Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426,** 442–446 (2003).
7. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci. e* **2,** 1527–1537 (1982).
8. Nieder, A., Freedman, D. J. & Miller, E. K. Representation of the quantity of visual items in the primate prefrontal cortex. *Science* **297,** 1708–1711 (2002).
9. Nieder, A. & Miller, E. K. A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences* **101,** 7457–7462 (2004).
10. Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience* **9,** 545–556 (2008).
11. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology* **37,** 66–74 (2016).
12. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497,** 585–590 (2013).
13. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nature Neuroscience* **19,** 973–980 (2016).
14. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441,** 223–226 (2006).
15. Kennerley, S., Dahmubed, A., Lara, A. & Wallis, J.D. Neurons in the frontal lobe encode the value of multiple decision variables. *J Cogn Neurosci. 21(6): 1162-78* (2009).
16. Cai, X. & Padoa-Schioppa, C. Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron* **81,** 1140–1151 (2014).
17. Padoa-Schioppa, C. Neuronal Origins of Choice Variability in Economic Decisions. *Neuron* **80,** 1322–1336 (2013).
18. Cai, X. & Padoa-Schioppa, C. Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J. Neurosci.* **32,** 3791–3808 (2012).
19. Raghuraman, A. P. & Padoa-Schioppa, C. Integration of multiple determinants in the neuronal computation of economic values. *J. Neurosci.* **34,** 11583–11603 (2014).
20. Hayden, B. Y. & Platt, M. L. Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* **30,** 3339–3346 (2010).
21. Lau, B. & Glimcher, P. W. Value Representations in the Primate Striatum during Matching Behavior. *Neuron* **58,** 451–463 (2008).
22. O'Neill, M. & Schultz, W. Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron* **68,** 789–800 (2010).
23. Padoa-Schioppa, C. & Assad, J. A. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience* **11,** 95–102 (2008).
24. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature Neuroscience* **10,** 1214–1221 (2007).
25. Critchley, H. D., Mathias, C. J. & Dolan, R. J. Neural Activity in the Human Brain Relating to Uncertainty and Arousal during Anticipation. *Neuron* **29,** 537–545 (2001).
26. McCoy, A. N. & Platt, M. L. Risk-sensitive neurons in macaque posterior cingulate cortex. *Nature Neuroscience* **8,** 1220–1227 (2005).
27. Ogawa, M. *et al.* Risk-responsive orbitofrontal neurons track acquired salience. *Neuron* **77,** 251–258 (2013).
28. Rushworth, M. F. S. & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience* **11,** 389–397 (2008).
29. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and

behavioural impact of decision confidence. *Nature* **455,** 227–231 (2008).
30. Lak, A. *et al.* Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. *Neuron* **84,** 190–201 (2014).
31. Meyniel, F., Sigman, M. & Mainen, Z. F. Confidence as Bayesian Probability: From Neural Origins to Behavior. Neuron **88,** 78–92 (2015).
32. Howard, J. D., Gottfried, J. A., Tobler, P. N. & Kahnt, T. Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proceedings of the National Academy of Sciences* **112,** 5195–5200 (2015).
33. Rudebeck, P. H. & Murray, E. A. The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron* **84,** 1143–1156 (2014).
34. Walton, M. E., Behrens, T. E. J., Buckley, M. J., Rudebeck, P. H. & Rushworth, M. F. S. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* **65,** 927–939 (2010).
35. Takahashi, Y. K. *et al.* Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron* **80,** 507–518 (2013).
36. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural mechanisms of foraging. *Science* **336,** 95–98 (2012).
37. Kennerley, S. W., Behrens, T. E. J. & Wallis, J. D. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience* **14,** 1581–1589 (2011).
38. Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J. & Rushworth, M. F. S. Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience* **9,** 940–947 (2006).
39. Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. Probabilistic brains: knowns and unknowns. *Nature Neuroscience* **16,** 1170–1178 (2013).
40. van Duuren, E. *et al.* Neural coding of reward magnitude in the orbitofrontal cortex of the rat during a five-odor olfactory discrimination task. *Learn. Mem.* **14,** 446–456 (2007).
41. van Duuren, E. *et al.* Single-Cell and Population Coding of Expected Reward Probability in the Orbitofrontal Cortex of the Rat. *J. Neurosci.* **29,** 8965–8976 (2009).
42. Deneve, S., Latham, P. E. & Pouget, A. Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience* **2,** 740–745 (1999).
43. Strait, C. E., Blanchard, T. C. & Hayden, B. Y. Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron* **82,** 1357–1366 (2014).
44. Rustichini, A. & Padoa-Schioppa, C. A neuro-computational model of economic decisions. *J. Neurophysiol.* **114,** 1382–1398 (2015).
45. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61,** 331–349 (1989).
46. Heekeren, H. R., Marrett, S., Bandettini, P. A. & Ungerleider, L. G. A general mechanism for perceptual decision-making in the human brain. *Nature* **431,** 859–862 (2004).
47. Malalasekera, W. Neuronal Mechanisms of Decision Making in the Prefrontal Cortex. *Doctoral thesis, UCL (University College London).* (2016).
48. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. in (1992).
49. Chang, C.-C. & Lin, C.-J. LIBSVM. *ACM Transactions on Intelligent Systems and Technology* **2,** 1–27 (2011).
50. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences* **101,** 13124–13131 (2004).
51. Stalnaker, T. A., Cooch, N. K. & Schoenbaum, G. What the orbitofrontal cortex does not do. *Nature Neuroscience* **18,** 620–627 (2015).
52. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron* **81,** 267–279 (2014).
53. Ma, W. J. & Jazayeri, M. Neural coding of uncertainty and probability. *Annual Review of Neuroscience* **37,** 205–220 (2014).

54. van Bergen, R. S., Ma, W. J., Pratte, M. S. & Jehee, J. F. M. Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience* **18,** 1728–1730 (2015).
55. Beck, J. M. *et al.* Probabilistic population codes for Bayesian decision making. *Neuron* **60,** 1142–1152 (2008).
56. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27,** 8366–8377 (2007).
57. Behrens, T. E. J., Hunt, L. T., Woolrich, M. W. & Rushworth, M. F. S. Associative learning of social value. *Nature* **456,** 245–249 (2008).
58. Hanks, T. D. *et al.* Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* **520,** 220–223 (2015).
59. Hayden, B. Y., Pearson, J. M. & Platt, M. L. Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience* **14,** 933–939 (2011).
60. Kolling, N., Behrens, T., Wittmann, M. K. & Rushworth, M. Multiple signals in anterior cingulate cortex. *Current Opinion in Neurobiology* **37,** 36–43 (2016).
61. Shenhav, A., Straccia, M. A., Cohen, J. D. & Botvinick, M. M. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience* **17,** 1249–1254 (2014).
62. Chan, S. C. Y., Niv, Y. & Norman, K. A. A Probability Distribution over Latent Causes, in the Orbitofrontal Cortex. *J. Neurosci.* **36,** 7817–7828 (2016).
63. Schoenbaum, G., Takahashi, Y., Liu, T.-L. & McDannald, M. A. Does the orbitofrontal cortex signal value? *Annals of the New York Academy of Sciences* **1239,** 87–99 (2011).
64. Padoa-Schioppa, C. & Schoenbaum, G. Dialogue on economic choice, learning theory, and neuronal representations. *Current Opinion in Behavioral Sciences* **5,** 16–23 (2015).
65. Lopatina, N. *et al.* Lateral orbitofrontal neurons acquire responses to upshifted, downshifted, or blocked cues during unblocking. *eLife Sciences* **4,** (2015).
66. Stalnaker, T. A. *et al.* Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nat Comms* **5,** 3926 (2014).
67. Tsutsui, K.-I., Grabenhorst, F., Kobayashi, S. & Schultz, W. A dynamic code for economic object valuation in prefrontal cortex neurons. *Nat Comms* **7,** 12554 (2016).
68. Fiser, J., Berkes, P., Orbán, G. & Lengyel, M. Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci. (Regul. Ed.)* **14,** 119–130 (2010).
69. deCharms, R. C. & Zador, A. Neural representation and the cortical code. *Annu. Rev. Neurosci.* **23,** 613–647 (2000).
70. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27,** 9984–9988 (2007).
71. Fitzgerald, T. H. B., Friston, K. J. & dolan, R. J. Action-Specific Value Signals in Reward-Related Regions of the Human Brain. *J. Neurosci.* **32,** 16417–16423 (2012).
72. Chib, V. S., Rangel, A., Shimojo, S. & O'Doherty, J. P. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* **29,** 12315–12320 (2009).
73. Gross, J. *et al.* Value Signals in the Prefrontal Cortex Predict Individual Preferences across Reward Categories. *J. Neurosci.* **34,** 7580–7586 (2014).
74. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76,** 412–427 (2013).
75. Boorman, E. D., Rushworth, M. F. & Behrens, T. E. Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. *J. Neurosci.* **33,** 2242–2253 (2013).
76. Lebreton, M., Jorge, S., Michel, V., Thirion, B. & Pessiglione, M. An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* **64,** 431–439 (2009).
77. Grueschow, M., Polania, R., Hare, T. A. & Ruff, C. C. Automatic versus Choice-Dependent Value

Representations in the Human Brain. *Neuron* **85,** 874–885 (2015).
78. Bendor, D. & Wang, X. The neuronal representation of pitch in primate auditory cortex. *Nature* **436,** 1161–1165 (2005).
79. Moser, E. I. *et al.* Grid cells and cortical representation. *Nature Reviews Neuroscience* **15,** 466–481 (2014).
80. Dordek, Y., Soudry, D., Meir, R. & Derdikman, D. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *eLife Sciences* **5,** e10094 (2016).
81. Louie, K., Grattan, L. E. & Glimcher, P. W. Reward Value-Based Gain Control: Divisive Normalization in Parietal Cortex. *J. Neurosci.* **31,** 10627–10639 (2011).
82. Polania, R., Krajbich, I., Grueschow, M. & Ruff, C. C. Neural Oscillations and Synchronization Differentially Support Evidence Accumulation in Perceptual and Value-Based Decision Making. *Neuron* **82,** 709–720 (2014).
83. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503,** 78–84 (2013).
84. Bush, D., Barry, C., Manson, D. & Burgess, N. Using Grid Cells for Navigation. *Neuron* **87,** 507–520 (2015).
85. Barlow, H. B. Possible principles underlying the transformations of sensory messages. *In Sensory Communication (1961), pp. 217-234* 217–234 (1961).
86. Simoncelli, E. P. Vision and the statistics of the visual environment. *Current Opinion in Neurobiology* **13,** 144–149 (2003).
87. Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518,** 529–533 (2015).
88. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352,** 1464–1468 (2016).
89. Stachenfeld, K. L., Botvinick, M. & Gershman, S. J. Design Principles of the Hippocampal Cognitive Map. *Advances in Neural Information Processing Systems* 2528–2536 (2014).
90. Louie, K. & Glimcher, P. W. Efficient coding and the neural representation of value. *Annals of the New York Academy of Sciences* **1251,** 13–32 (2012).

Chapter 7: General discussion

Our lives are delineated by the decisions that we make and coloured by the emotions that result. In this thesis I have attempted to shed light upon the way that different emotions arise, by characterizing them as the result of learning processes occurring in the brain (**Chapters 2** and **4**). I have also described the feedback influence of emotions upon subsequent learning, detailing a subtle deficit in action learning following stress (**Chapter 3**). In the final two experiments I changed tack, focusing upon the representation of value, central to the models of learning and decision-making discussed in the preceding chapters. We saw how fMRI can be used to assess the construction of value at the level of the whole brain, documenting the role of the anterior cingulate in representing value estimates composed of constituents that appeared to be separately represented elsewhere in the brain (**Chapter 5**). Focusing in on the anterior cingulate and the orbitofrontal cortex, I then provided evidence that the way that neurons represent value might not be the same as the way that humans think about value, as a linear quantity (**Chapter 6**). As each chapter contains a relatively extensive discussion of the issues pertinent to that study, in this summary I will try and draw together some common threads and point out where there are holes in the patchwork.

7.1 Optimality's ever widening net

The work presented here that deals with emotion is founded upon the idea that emotion is predictable and useful – that there is a logic behind fluctuations in happiness and stress that is expressed in their causes and consequences. This is a relatively recent perspective, reflecting the increasingly common assumption that the brain is best understood as a constrained solver of computational problems. In this case, we have touched upon the idea that mood might relate to trajectories of reward in the environment, and that stress might encourage us to avoid situations of great uncertainty. Another possibility which we have not discussed is that emotions can play a fundamentally social role, providing a convenient way to broadcast information about the environment and an organism's interaction with it. In any case, we have come a long way from the Enlightenment notion of emotion as an impediment to clear thinking.

Similarly, the argument for probabilistic population codes put forward in **Chapters 5** and **6** stems from an analysis of how noisy information channels can best convey information consistent with Bayesian computation. This comes on the back of a decade or so of work establishing Bayes-optimal, or near-optimal, behaviour in a wide variety of tasks. Similarly, a great deal of effort has been expended detailing the rationality of choice and our deviations from it. As we shall discuss below, this emphasis on optimal solutions requires us to think very carefully about how our problems are formulated. The description of the problem defines the optimal solution; recent work has emphasised how optimal solutions can appear sub-optimal if the problem is inadequately characterized [1]. Arguably, this has been an issue in the study of value-based decision-making: we have focused upon optimal solutions to isolated gambling problems without first asking whether these kind of choices are representative of real decisions.

7.2 Value: economics, nature, and neural networks

In the early sections of this thesis I described classic formulations of value, such as those undertaken by Pascal, Bernoulli, and others of an economic bent. The ideas of economics have had a profound impact upon the study of decision-making in neuroscience. The allure of economic descriptions is that they condense behaviour into a few clean numbers; their failing is that they provide a wholly unrealistic description of the problems that the brain has evolved to solve, and pay no attention to how an organism *learns* to solve those problems. One unhelpful mental model that we have inherited from economics is the idea that value is a monotonic quantity. Not only is this demonstrably untrue in behavioural terms, with humans routinely violating transitivity [1,2], but, as I have argued, it seems like an unhelpful assumption when considering the neural representations underlying value-based choice.

As discussed in **Chapter 6**, a more reasonable approach might be to consider the kinds of problems the brain is trying to solve. In this view, the representation within the brain is always subservient to the computation the brain is trying to achieve, a philosophy naturally expounded by Marr's three levels of analysis [3]. Marr argued that we can understand an information in terms of computation ('what is it trying to do'), algorithm ('how the software work?'), and implementation ('how does the hardware work?'). Through this lens, we can see that the idea that value representations ought to be linear results from a cross-contamination of

characterizations at the algorithmic and implementational levels: the assumption of linearity has leaked from software to hardware.

Starting with the computational problem often dictates understanding the features of the problem as it typically occurs in nature. This approach has yielded great insight in the sensory domain, with an analysis of the natural statistics of images providing an excellent description of tuning properties in early visual cortex [4]. As the problems we study become more remote from the sensory information on which they are based, this becomes more difficult. Not only are the necessary statistics harder to define - what are the natural statistics of reward? – but the current statistics might not reflect those that drove the evolution of the brain. A great deal of variance in dopaminergic activity today might be driven by Facebook likes, but understanding how the brain represents the 'likeability' of social media content is not a useful quest. Although such thinking has led to an increasing interest in foraging and time-limited problems [5,6], there remains much we don't know about how best to characterize choice in an ethological way, and whether there remains a role for value in that description [7].

Recent advances in neural networks provide another avenue through which to explore complex cognitive processes. Although there are problems associated with interpreting the relationship between brain data and models (see below), the hope is that by designing neural networks to do interesting value-based tasks, we can glean insight into how the brain achieves this and other feats of complex cognition [8]. Drawing again on Marr's three levels, we can see several ways in which neural networks might help us understand the brain. At an algorithmic level, neural networks might take forms which directly recapitulate the brain [9]. At a computational level, they might provide some hints as to the problem features or technical complexities that we should be thinking about to understand the brain. For instance, recent networks have made use of adversarial architectures, with the performance of one network sharpened by its attempt to fool another [10]. Arguably, this mirrors the competitive context in which brains develop, a topic which has received little attention in value-based decision-making. Nearly all of the decisions that an animal makes involve deciding whether it's worth *competing* for a reward, be it sex, food, or shelter. This is particularly true in the complex social hierarchies in which the human brain evolved.

However, the analogy between synthetic and natural neural networks can only be stretched so far. It is important to acknowledge that *constraints* play a crucial role in the architecture of neural circuits, and the constraints faced by synthetic and real neural networks are likely to be very different. Spatiotemporal [11] and energetic [12] considerations are known to play a profound role in the design and function of real neural networks. Conversely, much energy is expended in the design of artificial neural networks to format them in a manner appropriate for GPU processing [13]. The point at which the analogy between artificial and natural neural networks breaks down due to these differences remains to be seen.

7.3 Models as bridges and sirens

A dominant theme in this thesis has been that of models. Computational models provide us with the ability to make precise predictions about phenomena, bringing a quantitative bearing upon things that might at first sight not appear amenable to scientific enquiry, such as emotion. They can therefore offer us the ability to link disparate domains; as we saw in **Chapters 2** and **4**, models originally constructed to explain learning can be co-opted to describe emotional dynamics. Similarly, we can use models to bind together measurements, such as subjective and physiological assays of stress, by characterizing their dependence upon variables in our model. We can further reassure ourselves by fitting our models to separate individuals, and checking that interindividual variability in different measurement modalities is coherent. Models are indispensable to modern neuroscience, and rightly so.

However, models also have the power to beguile. By bringing to bear machinery from disciplines unfamiliar to cognitive scientists, models can tempt us to characterize our data in a way that is overcomplicated, or obfuscatory. More worryingly, models can blind us to questions of specificity. For example, the first model of repetition suppression reported in **Chapter 5** supported the idea put forward in the literature, and justified by similar models, that repetition suppression signals of the type we observe are evidence of Gaussian tuning. However, closer investigation showed that these conclusions were contingent upon decisions made during model implementation that might not be immediately apparent to the reader. One suspects that this is frequently the case in cognitive neuroscience. Consequently, use of models to support interpretation of data might be considered proof of plausibility rather than empirical proof.

Additionally – and this is something of a marmite quality- models also force us to ask what we mean when we say we *understand* something. For instance, in **Chapter 4** we show that a Bayesian hierarchical model provides a good description of learning in unpredictable environments. How much faith does this give us that the brain implements a hierarchical Gaussian filter? The problem might seem to be alleviated by collecting brain data, but difficulties persist. If fMRI signal – or neural firing rates – in some part of the brain bears a resemblance to either the predictions or the hidden variables in the HGF, does that constitute *proof*? Phrased another way, it is usually challenging to design an algorithm to solve the task at hand that displayed *no* relationship to behaviour and the brain. This has been emphasised by recent reports that deep neural networks produce patterns of representational similarity that echo those observed in parts of the visual system [9,14,15]. By and large, the more a neural network's internal representations resemble those of the brain, the better job they do of classifying images. Optimistically, one would say that we have discovered evidence that the brain uses deep neural networks to recognize images. Pessimistically, one would counter that we've merely shown that informational processes that do a good job of turning images into certain labels pass through similar intermediate states, a worry expressed by the reticent title of a recent paper 'Deep Supervised, but Not Unsupervised, Models *May* Explain IT Cortical Representation' [emphasis mine] [15]. The promise of modelling is that it allows us to achieve a mechanistic description of brain activity: but in most cases, doesn't it really just allowing us to describe our model, describe the brain, and then draw parallels? As we become more ambitious in our attempts to understand the brain, we may have to remind ourselves of George Box's classic maxim – 'all models are wrong, but some models are useful' – and consider how we should assess the usefulness of a model in neuroscience.

7.4 Concluding remarks

The picture that the work in this thesis paints of the brain is one of numerous feedback loops, acting in parallel. I have emphasised how emotion provides both an input and an output to learning and how value might be assembled and represented in the cortex in a format compatible with rapid updating. Although all the work reported here took part in a trial-based fashion, my hope is that future work develops more sophisticated models of learning and emotion with dynamics that more closely resemble those that we experience on a day to day

basis. It will be fascinating to see whether our models will withstand the rough and tumble of realistic environments with naturally behaving agents.

This research suggests promising routes to understanding slower fluctuations in emotion – such as life satisfaction or chronic stress- in relation to the faster processes characterised here. Given the importance of these variables to long-term health and the ever-increasing burden imposed by psychiatric disorders of emotion, the importance of this mission can hardly be overstated. Assessing whether the computational models that usefully characterize emotions in the laboratory will prove a clinically relevant tool is an exciting and pressing challenge.

7.5 References

1. Summerfield, C. & Tsetsos, K. Do humans make good decisions? *Trends Cogn. Sci. (Regul. Ed.)* **19,** 27–34 (2015).
2. Tsetsos, K. *et al.* Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences* **113,** 3102–3107 (2016).
3. Marr, D. Vision: A computational investigation into the human representation and processing of visual information. (1982).
4. Barlow, H. B. Possible principles underlying the transformations of sensory messages. *In Sensory Communication* 217–234 (1961).
5. Kolling, N., Behrens, T. E. J., Mars, R. B. & Rushworth, M. F. S. Neural mechanisms of foraging. *Science* **336,** 95–98 (2012).
6. Kolling, N., Wittmann, M. & Rushworth, M. F. S. Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* **81,** 1190–1202 (2014).
7. Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience* **33,** 269–298 (2010).
8. Marblestone, A. H. & Wayne, G. Toward an integration of deep learning and neuroscience. *Front. Comput. Neurosci.* (2016).
9. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience* **19,** 356–365 (2016).
10. Goodfellow, I. *et al.* Generative Adversarial Nets. *Advances in Neural Information Processing Systems* 2672–2680 (2014).
11. Rivera-Alba, M. *et al.* Wiring Economy and Volume Exclusion Determine Neuronal Placement in the Drosophila Brain. *Current Biology* **21,** 2000–2005 (2011).
12. Attwell, D. & Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21,** 1133–1145 (2001).
13. Schmidhuber, J. Deep learning in neural networks: an overview. *Neural networks : the official journal of the International Neural Network Society* **61,** 85–117 (2015).
14. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* **111,** 8619–8624 (2014).
15. Khaligh-Razavi, S.-M. & Kriegeskorte, N. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comp Biol* **10,** e1003915 (2014).